

Acquisition of clinical competence during postgraduate training in general practice

Citation for published version (APA):

Kramer, A. (2003). *Acquisition of clinical competence during postgraduate training in general practice*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20030611ak>

Document status and date:

Published: 01/01/2003

DOI:

[10.26481/dis.20030611ak](https://doi.org/10.26481/dis.20030611ak)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

ACQUISITION OF CLINICAL
COMPETENCE DURING
POSTGRADUATE TRAINING
IN GENERAL PRACTICE

VERWERVING VAN KLINISCHE
COMPETENTIE TIJDENS DE
HUISARTSOPLEIDING

Anneke Kramer

Cover Design	Maarten Kramer
Design and lay-out	MTM Grafische Studio, UMC Utrecht
Pictures	Maarten Kramer
Printing	Datawyse / Universitaire Pers Maastricht

ISBN	90-5394-097-9
©	Oxford University Press, Chapter 1
©	Blackwell Publishing Ltd, Chapters 2 and 3
©	Royal College of General Practitioners, Chapter 4
©	Anneke Kramer, Other chapters

ACQUISITION OF CLINICAL COMPETENCE DURING POSTGRADUATE TRAINING IN GENERAL PRACTICE

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof.dr. A.C. Nieuwenhuijzen Kruseman
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 11 juni 2003 om 16.00 uur

door

Anneke Kramer

Promotores:

Prof.dr. C.P.M. van der Vleuten

Prof.dr. R.P.T.M. Grol

Co-promotor:

Dr. L.H.C. Tan

Beoordelingscommissie :

Prof.dr. J.F.M. Metsemakers (voorzitter)

Dr. J. Cohen-Schotanus (Rijksuniversiteit Groningen)

Dr. Y. van Leeuwen

Prof.dr. E. Schadé (Universiteit van Amsterdam)

Prof.dr. A.J.J.A. Scherpbier

This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

The SVUH provided financial support for the publication of this thesis.

Aan mijn ouders

In dierbare herinnering aan mijn oma (1903-2002)

Houd Ithaka wel altijd in gedachten.
Daar aan te komen is je doel.
Maar overhaast de reis in geen geval.
Beter is dat die vele jaren duurt,
zodat je oud zult zijn wanneer je bij het eiland
het anker uitwerpt, rijk aan wat je onderweg verwierf,
en niet verwachtend dat Ithaka je rijkdom schenken zal.
Ithaka gaf je de mooie reis.
Was het er niet, dan was je nooit vertrokken,
verder heeft het je niets te bieden meer.

En vind je het wat pover, Ithaka bedroog je niet.
Zo wijs geworden, met zoveel ervaring, zul je al
begrepen hebben wat de Ithaka's beduiden.

Uit Ithaka van K.P. Kavafis, in de vertaling van Hans Warren en Mario Molengraaf

Contents

Introduction	01
Chapter 1	
Effect of extension of postgraduate training in general practice on the acquisition of knowledge of trainees <i>Family Practice 2003;20:207-212</i>	13
Chapter 2	
Comparison of a rational and an empirical standard setting procedure for an OSCE <i>Medical Education 2003;37:132-139</i>	27
Chapter 3	
Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice <i>Medical Education 2002;36:812-819</i>	45
Chapter 4	
Acquisition of clinical skills in postgraduate training for general practice <i>Accepted for publication in the British Journal of General Practice</i>	59



Chapter 5	
Acquisition of communication skills in postgraduate training for general practice	75
Chapter 6	
Growth of self-perceived clinical competence in postgraduate training for general practice	91
General conclusions and discussion	105
Summary of this thesis	123
Samenvatting van dit proefschrift	137
Dankwoord	151
Curriculum vitae	157





INTRODUCTION

The purpose of this study is to investigate the effects of the proposed system on the performance of the system. The results of the study are presented in the following sections.

This thesis describes an evaluation study into the effects of postgraduate training in general practice on the development of clinical competence of trainees. This chapter provides the rationale for focusing on this area; the context of the study; the general aims and specific research questions; the outcome measures and general design; and the outline of the thesis.

Rationale for evaluating postgraduate training in general practice

Due to an increasing demand and growing complexity of the provision of primary health care,^{1,2} many countries have, over the last 30 years, introduced postgraduate training programmes for general practice.^{2,8} Surprisingly, even between countries with comparable health care systems, and thus with similar demands on the role of the general practitioner (GP) therein, a considerable variation exists in structure, content and duration of these programmes.^{2,8} This can be taken to mean that different approaches could apply, all resulting in the same qualified GP, in the sense that many ways lead to Rome. However, few studies have been published that report on the effectiveness of these different programmes. In fact, despite the many efforts to provide a comprehensive education, little evidence exists whether or not Rome is actually reached. In educational terms, we lack proper evidence, whether or not sufficient numbers of trainees, at completion of their training, have attained the learning objectives of the programme.^{6,9,10} There are good reasons to perform such evaluation. One is the increasingly important and complex role of the GP in the provision of primary health care and this requires adequate and successful training.² The other reason is of an economical order, the expenses of postgraduate training have to be justified.¹¹⁻¹³ To give an idea, the budget for Dutch postgraduate training in general practice amounted to 62 million Euros in 2002. Finally, in line with the current development of evidence-based medical practice, one could envision the development of evidence-based medical education.¹¹⁻¹⁹ Such development is driven by a continuous evaluation of the effects of medical education programmes and it is this evaluation that is at the heart of this thesis. Here we put forward a number of parameters to measure the effects of education and we provide the outcome of a number of such analyses.

Context of the study

History

Postgraduate training in general practice in The Netherlands is in a continuous process of change. Until 1973, it was possible to start in a general practice immediately after graduating in medicine.

Professional developments in general practice instigated the need for a specific training programme and this resulted in the provision, in 1973, of a compulsory one-year postgraduate training programme, provided by the departments of general practice of all Dutch medical schools (eight). The publication, in 1983, of the Basic Job Description for the GP by the Dutch National Association of General Practitioners (LHV) initiated a further elaboration of the postgraduate training.²⁰ The committee 'Construction of a new Curriculum of Postgraduate Training for General Practice' (CCBOH) proposed a three-year curriculum, based on this Job Description.²¹ In first instance, the government consented to an extension to only two years. This two-year interim programme started in 1988. Subsequent evaluation of the content of this programme demonstrated shortcomings and this, in 1994, opened the way for a further extension to three years.^{22,23} In 2001, a new blueprint for the curriculum was drawn up again to accommodate novel teaching methods (such as self-directed learning and modular teaching) and to accommodate altered learning objectives.²⁴ This curriculum is to be implemented in the coming years.

The study presented in this thesis concerns an evaluation of the two-year and the three-year curriculum up to and including 1999.

Training Programme

Based on the Basic Job Description for the GP, the committee CCBOH made proposals for the structure and design of the curriculum, its learning objectives, the tasks of the trainers and the evaluation of trainees and the programme itself. The eight institutes for postgraduate training in general practice accepted these proposals and this provided, from 1988 onwards, a greater degree of uniformity between their training programmes.

The curriculum had to offer training for the following tasks (and to be mastered by the trainees):

1. Tasks related to the process of providing care, problem solving and care provision adapted to the situation in general practice;

2. Tasks related to certain categories of patients, complaints and diseases;
3. Supportive tasks: ability to organise and manage a general practice; to cooperate with other care providers; and to maintain professional competence;
4. Tasks related to the personal functioning as a GP.

Concrete learning objectives were formulated for each of these tasks, specified in three subgroups related to the following aspects of clinical competence: knowledge, skills (including clinical skills and communication skills) and attitude.

Apart from what had to be learned, the committee CCBOH also identified where and from whom this had to be learned. Four learning environments were recognised:

1. Learning in the general practice;
2. Learning in other clinical settings (e.g. hospitals, outpatients clinics);
3. Learning at the training institute;
4. Self-directed learning.

Furthermore, the committee CCBOH defined the structure of the curriculum, proposing a curriculum divided into three blocks of equal length, each with specified learning objectives and a specific learning environment. In the two-year curriculum (1988-1994), each block took about eight months. In the three-year curriculum, these blocks are one year each. The first and third blocks are reserved for practical training in a general practice and the second block is dedicated to rotations in hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. The second block is deliberately placed after the first, to ensure that this training environment is perceived from the learning perspective of a GP-trainee. Throughout training in a general practice, each trainee is coached by a GP-trainer (an experienced GP who has attended training for supervision and is recognised by an independent legal body). The GP-trainer of the first year differs from the one of the third year and, within practical constraints, the kind of practice differs as well (for example rural against urban). The learning objectives of the first block are to gain knowledge, skills and attitudes in the field of general practice with emphasis on common complaints, disorders and problems. The aim of the second block is the acquisition of additional knowledge, skills and attitudes relevant to general practice that cannot be learned satisfactorily in general practice itself. Finally, in the third block trainees have to integrate their new knowledge, skills and

attitudes into their activities as a GP with emphasis on management of complex situations like preventive and continuous care. They also have to prepare themselves for independent practice. To enable the practical training to be a real learning situation, trainees are, in economical sense, 'unproductive': their apprenticeship does not provide additional income for the host institute.

In addition to the practical training, trainees attend one day per week meetings, in groups of about 12, at the training institute for education in theory, for clinical and communication skills training and for reflection. A half-day per week is reserved for self-directed learning.

The content, structure and learning objectives of the curriculum are briefly outlined in Box 1.

Box 1

Postgraduate training in general practice: outline of the curriculum at the time of the study (1992-1999)

CONTENT			
Basic Job Description for the general practitioner			
STRUCTURE	Three blocks of equal length		Throughout the curriculum
	Block 1	General practice training	One day per week for; Half-day per week for;
	Block 2	Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics	special training and reflection, in groups of trainees, at the training institute; self-directed learning
	Block 3	General practice training	
LEARNING OBJECTIVES	Block 1	Acquisition of knowledge, skills and attitudes with emphasis on common problems	
	Block 2	Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself	
	Block 3	Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations	

Evaluation

Trainee evaluation

For the evaluation of the trainees, a distinction was made between educational assessment, aimed at supporting the learning process of the trainee, and selective assessment, aimed at selection and certification. The first year of training is most important, because trainees have to pass the progress qualification at the end of this year in order to be allowed to continue postgraduate training.²⁵ During this year trainees receive feedback about their learning process and progress from both their GP-trainer and their teachers at the training institute. The evaluation process is based on subjective ratings of the performance on the four main tasks outlined above, and on written knowledge tests. At the end of the first year, GP-trainers and teachers complete an advisory form for the purpose of the progress qualification. They are obliged to include the results of two written tests and the subjective ratings. Additional information may be provided by the submission of performance indicators like direct or audio/video taped observations of consultations, judgement of patients, judgement of colleagues or other personnel of the general practice, results of performance based tests and, lastly, the contribution of the trainee to the group meetings at the training institute. Based on the progress qualification report, the head of the training institute decides whether a trainee may continue postgraduate training, needs remediation or has to stop. If a trainee has passed the progress qualification and has completed the three-year of training, he/she will be registered as a GP and is allowed to commence general practice independently. Only when serious doubts exist about the abilities of a trainee after passing the progress qualification, the head of the training institute may decide to advise remediation or disqualification. The progress qualification procedure is based on legal guidelines and has to be applied uniformly by the eight training institutes.²⁶ After the first year, GP-trainers and teachers continue to assess trainees on an educational basis, using subjective ratings as well as objective measures of clinical competence. The extent to which and how this assessment takes place, is the responsibility of the individual training institutes.

Programme evaluation

The evaluation of the training programme itself was not guided by national procedures. During the years national and local evaluation

programmes have been performed. However, training institutes, training clinics and GP-trainers have to be recognised by an independent legal body, the Registration Committee for Postgraduate Training in General Practice/Family Medicine (HVRC). At least once in the five years the HVRC evaluates whether extension of the recognition is justified according to several quality criteria.

National evaluation

Since 1987 the eight training institutes cooperate in a National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).²⁷ This centre has two main tasks: (1) to develop, produce and administer competence tests; and (2) to develop and perform national evaluation studies. From 1995 to 1999 a comprehensive national evaluation (EVA-H project) took place,²⁸ including:

1. The competence of trainees: do they master the tasks of the GP as defined by the Basic Job Description at completion of a three-year training programme?
2. The structure and content of the curriculum: do the training institutes organise the curriculum according to the guidelines of the committee CCBOH?
3. A comparison between the level of competence of the two-year and the three-year curriculum: does the three-year curriculum lead to a higher level of competence?

General aims and specific research questions of the study

General aims

The study presented in this thesis was part of the EVA-H project mentioned in the paragraph above. It focused on the outcome and comprised the following questions: to what extent do trainees, at completion of their training, master the tasks of the GP as defined by the Basic Job Description; and is there any difference between the two- and the three-year curriculum? Apart from examining the final outcome, we were also interested in the process of acquisition of clinical competence during the course and to what extent postgraduate training contributed to that. Was there a significant and meaningful increase from start to finish? Was the increase a rather continuous process or were there particular parts of the curriculum that contributes more or less?

This information enabled us to evaluate the effectiveness of the training programme. We have limited the study to the evaluation of task 1, tasks related to the process of providing care, and task 2, tasks related to certain categories of patients, complaints and diseases. Task 3, supportive tasks, and task 4, tasks related to the personal functioning as a GP, were thought to be too complex for national evaluation. Moreover, we only assessed learning objectives related to the domains knowledge and skills. Attitude was not included for the same reason as tasks 3 and 4 were excluded.

Finally, this thesis also presents some methodological studies, in order to facilitate answering of the main research questions. Firstly, we needed standards of adequacy to decide whether a satisfactory level was achieved at completion of training or not. Standard setting is a rather judgmental process, involving an arbitrary decision of what is considered to be 'good enough'. To guarantee that this process is defensible and controllable, aspects of credibility and reliability have to be considered. In our study we focused on a standard setting procedure for a performance-based test of clinical skills. Secondly, we required an assessment instrument for large-scale evaluation of clinical skills. Although the Objective Structured Clinical Examination (OSCE) is one of the most valid methods for the assessment of clinical skills, broad application is complex and expensive. As a knowledge test is more feasible to apply, we were interested in to what extent a knowledge test of skills could predict performance on an OSCE.

Specific research questions

Based on the general aims of the study the following specific research questions were defined:

1. What is the acquisition of knowledge during a two-year and three-year postgraduate training programme for general practice, and does a three-year programme lead to a higher knowledge level than a two-year programme, while maintaining the same structure and learning objectives?
2. What are the credibility and reliability of a rational and an empirical standard setting procedure in an OSCE, and what is the impact of a reality check on the credibility and reliability of the rational procedure?
3. What is the predictive validity of a Knowledge Test of Skills for performance on an OSCE in postgraduate training for general practice?

4. What is the acquisition of clinical skills during a three-year postgraduate training programme for general practice, and is a satisfactory level achieved at completion of training?
5. What is the acquisition of communication skills during a three-year postgraduate training programme for general practice, and is a satisfactory level achieved at completion of training?
6. What is the increase of self-perceived clinical competence of trainees in a three-year postgraduate training programme for general practice, and how can the growth of self-perceived clinical competence be explained?

Outcome measures and general design of the study

Outcome measures

Having determined the research questions of the study, measurement instruments had to be chosen. A number of criteria exists to determine the usefulness of instruments.^{13,15,29-33} These criteria are: feasibility, acceptability, costs, reliability, validity and educational consequences. Considering these criteria, the following choices were made. As this study focused on national assessment of trainees, feasibility was a very important condition, followed by acceptability and costs to guarantee feasibility. Another requirement was the use of instruments with known levels of validity and reliability to assure valid and reliable results. Consequently, instruments that were already applied in postgraduate training were chosen. Additional advantages of this choice were that the trainees were accustomed to the assessment procedures and that the teaching procedures and learning objectives of the training programme were matched. Furthermore, it was thought to be essential to use a mix of different instruments in order to get a more comprehensive impression of the knowledge and skills of trainees. As case-specificity threatens a valid and reliable assessment of clinical competence, an additional advantage of this approach was that it enabled broad sampling of cases and problems.³⁴ According to the conceptual model of clinical assessment as devised by Miller, it was decided to evaluate all levels (knows, knows how, shows how and does).^{32,34} Finally, we were not only interested in the level of clinical competence assessed by objective measures but also in how trainees considered their own clinical competence, as this is an important guiding principle for self-directed learning and professional autonomy.

The choices mentioned above resulted in the use of the following instruments:

1. A general knowledge test with a progress-testing format (*knows and knows how*);⁹
2. A knowledge test of skills with a progress-testing format (*knows how*);³⁰
3. An OSCE (*shows how*);³⁰
4. A video-observation of real life consultations (*does*);³⁵
5. A self-assessment questionnaire.

General design

A retrospective design was adopted to compare the two-and three-year curriculum. As outcome measure the scores of general knowledge tests with a progress-testing format were used. Progress-testing implies that each test has the intention to reflect the cognitive final objectives of the curriculum.^{9,36} Such knowledge tests are periodically administered to all Dutch trainees in general practice at fixed intervals during the course to monitor progress. In a mixed-longitudinal design, both the acquisition of knowledge throughout training and the final result were examined.

For a more profound investigation of the three-year curriculum, a national cohort of all trainees starting in September 1995 was followed during three years. Various tests were administered to different samples of this cohort. The level of knowledge and skills at completion of training was measured using all instruments. For the acquisition of knowledge and skills throughout training, all instruments but the OSCE were used.

In order to judge whether a satisfactory level was achieved at completion of training, several procedures were applied: comparison with students (general knowledge), comparison with experienced GPs (knowledge of skills, OSCE, video-observation) and criterion-referenced standards (knowledge of skills, OSCE, video-observation).

Outline of the thesis

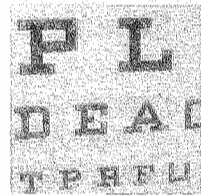
The studies concerning the six specific research questions are described in the chapters one to six. As each chapter is based on an article some repetition is inevitable. In the 'General conclusions and discussion' the main findings of the six studies are outlined and discussed, and recommendations for the future are given.

References

1. *Tarimo E, Webster EG.* Primary Health Care: Concepts and Challenges in a Changing World. Alma-Ata Revisited. 1-118. Geneva: WHO; 1997.
2. *Haq C, Ventres W, Hunt V, Mull D, Thompson R, Rivo M, Johnson P.* Where there is no family doctor: the development of family practice around the world. *Acad Med* 1995;70:370-80.
3. *Pabst R.* Medical education and reform initiatives in Germany. *Acad Med* 1995;70:1006-11.
4. *Campbell LM, Murray TS.* The effects of the introduction of a system of mandatory formative assessment for general practice trainees. *Med Educ* 1996;30:60-4.
5. *Cassie JM, Armbruster JS, Bowmer MI, Leach DC.* Accreditation of postgraduate medical education in the United States and Canada: a comparison of two systems. *Med Educ* 1999;33:493-8.
6. *Spike NA, Hays RB.* Analysis by training status of performance in the certification examination for Australian family doctors. *Med Educ* 1999;33:612-15.
7. *Campos-Outcalt D, Lundy M, Senf J.* Outcomes of combined internal medicine-pediatrics residency programs: a review of the literature. *Acad Med* 2002;77:247-56.
8. www.euract.org.
9. *Van Leeuwen YD.* Growth in knowledge of trainees in general practice [dissertation]. Maastricht: Universitaire Pers Maastricht; 1995.
10. *Irby D.* Teaching and learning in ambulatory care setting: a thematic review of the literature. *Acad Med* 1995;70:898-931.
11. *Murray E.* Teaching internal medicine in general practice: an evaluation [dissertation]. Maastricht: Department of Educational Development & Research; 2001.
12. *Remmen R.* An evaluation of clinical skills training at the medical school of the University of Antwerp [dissertation]. Antwerpen: Cuypers; 1999.
13. *Murray E.* Challenges in educational research. *Med Educ* 2002;36:110-12.
14. *Tan LHC.* Tekorten in de opleiding van huisartsen (Deficiencies in postgraduate training for general practice) [with English summary] [dissertation]. Amsterdam: Universiteit van Amsterdam; 1989.
15. *Bordage G, Burack JH, Irby DM, Stritter FT.* Education in ambulatory settings: developing valid measures of educational outcomes, and other research priorities. *Acad Med* 1998;73:743-50.
16. *Friedman CP, De Blik R, Greer DS et al.* Charting the winds of change: evaluating innovative medical curricula. *Acad Med* 1990;65:8-14.
17. *Harden RM, Grant J, Buckley G, Hart IR.* BEME Guide No.1: Best Evidence Medical Education. *Med Teacher* 1999;21:553-62.
18. *Prystowsky JB, Bordage G.* An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ* 2001;35:331-6.
19. *Van der Vleuten CPM, Dolmans DHJM, Scherpbier AJJA.* The need for evidence in education. *Med Teacher* 2000;22:246-50.

20. *Springer MP, editor.* Basic Job Description for the general practitioner. Utrecht: Dutch National Association of General Practitioners (LHV); 1983.
21. *Dubois V, Everwijn S, van Geldorp G et al.* The construction of a new curriculum of vocational training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
22. *Pollemans MC, Tan LHC.* Toetsing van kwaliteit (Assessing quality). Utrecht: SV-IOH; 1990.
23. *Wigersma L, Almekinders F, Kooij LR.* Raamplan curriculum driejarige huisartsopleiding (Blueprint curriculum three-year postgraduate training for general practice). Amsterdam/Utrecht: 1994.
24. *Onstenk JHAM, Swager R.* Raamplan Huisartsopleiding (Blueprint postgraduate training for general practice). 's Hertogenbosch: Cinop; 2001.
25. *Werkgroep kwaliteitsbewaking.* Kwaliteitsbewaking van de individuele huisarts-in-opleiding (Assessment of the individual general practice trainee). Utrecht: SVUH; 1995.
26. Addendum bij besluit CHVG NO.4-2000.
27. *Tan LHC, Carlier IVE.* Jaarverslag 2000 (Annual report 2000). Utrecht: SVUH; 2001.
28. *Tan LHC, Kramer AWM.* Opleiding doorgelicht (National evaluation study of Dutch postgraduate training in general practice). Utrecht: SVUH; 1999.
29. *Van der Vleuten CPM.* The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
30. *Jansen K.* Toetsing van technische vaardigheden van huisartsen (Assessment of technical skills of general practitioners) [with English summary] [dissertation]. Maastricht: Unigrafic; 1998.
31. *Wass V, van der Vleuten C, Shatzer J, Jones R.* Assessment of clinical competence. *Lancet* 2000;357:945-9.
32. *Miller GE.* The assessment of clinical skills/competence/performance. *Acad Med* 1990;65 (9Suppl):S63-7.
33. *Kassebaum DG, Eaglen RH.* Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med* 1999;74:842-9.
34. *Van der Vleuten.* Validity of final examinations in undergraduate medical training. *BMJ* 2000;321:1217-9.
35. *Ram P.* Comprehensive assessment of general practitioners [dissertation]. Maastricht: Unigrafic; 1998.
36. *Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW.* Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teacher* 1996;18:103-9.

Chapter 1



EFFECT OF EXTENSION OF POSTGRADUATE TRAINING IN GENERAL PRACTICE ON THE ACQUISITION OF KNOWLEDGE OF TRAINEES

This article has been published as:

Kramer AWM, Düsman H, Tan LHC, Jansen JJM, Grol RPTM and van der Vleuten CPM.
Effect of extension of postgraduate training in general practice on the acquisition of
knowledge of trainees. Family Practice 2003;20:207-212.

Summary

Purpose

Postgraduate training in general practice aims to develop clinical competence. However, little is known about its effect on trainees' development of competence. In this study is investigated whether three years of postgraduate training in general practice leads to a higher level of knowledge than two years of training, while maintaining the same structure and learning objectives.

Method

This retrospective study had a mixed longitudinal design. Trainees of the Dutch postgraduate training in general practice participated. Knowledge was assessed by written progress-testing of knowledge relevant to general practice embedded in real life situations. Test results were collected from 1992 to 1999. The results of trainees who received the two-year and three-year curriculum were compared.

Results

Both curricula started with the same entrance level and showed the largest acquisition of knowledge at the start and towards the end of training. Both curricula showed stagnation in growth at the end of the training period in which trainees rotate through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. The level of knowledge at the end of training was significantly higher for the three-year curriculum.

Conclusion

This study shows that postgraduate training contributes to an increase in knowledge and that a three-year programme leads to a higher level of knowledge than a two-year programme. Since knowledge is core to any kind of clinical competence, the stagnation in growth found at the end of rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics questions the impact of these rotations on the development of competence and the effectiveness of the training programme as a whole. Further study is needed to draw more firm conclusions.

Introduction

Postgraduate training in general practice shows considerable variation in duration and content, even between countries with comparable health care systems.¹⁻⁷ Although it has been demonstrated that general practitioners (GPs) with postgraduate training provide high quality and cost-effective care³ and, moreover, perform better during their career than those who did not receive such training,⁸ the effect of postgraduate training on the development of competence of trainees has only been evaluated rarely.^{9,10} One study into postgraduate training for psychiatrists demonstrated incremental growth in competence during their three-year programme.¹⁰ Yet, insight into the change in trainees' competence over time is valuable to evaluate postgraduate training. This is the focus of the study presented here. We investigated the effect of the extension of a postgraduate training programme in general practice on the acquisition of knowledge.

The content of postgraduate training in The Netherlands is based on the Basic Job Description for the GP.¹¹ The curriculum has a clearly outlined structure, being composed of three blocks of equal length, each with specified learning objectives (see Box 1).¹² The first and third blocks are reserved for practical training in a general practice, whereas the second block is dedicated to rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. This block is placed deliberately after the first to ensure that this training is perceived as a learning environment for a GP trainee rather than a specialist (hospital) training. The learning objectives of the first block are to gain knowledge, skills and attitudes in the field of general practice with emphasis on common complaints, disorders and problems. The goal of the second block is the acquisition of additional knowledge and skills relevant to general practice that cannot be learned satisfactorily in general practice itself. Finally, in the third block trainees have to integrate their new knowledge and skills into their activities as a GP with emphasis on management of complex situations such as preventive and continuous care. They also have to prepare themselves for independent practice. In addition to the practical training, trainees meet one day per week for education in theory and for reflection, and a half-day per week is reserved for self-directed learning.

Box 1

Dutch postgraduate training in general practice at the time of the study (1992-1999)

CONTENT				
Basic Job Description for the general practitioner				
STRUCTURE	Three blocks of equal length		Throughout the curriculum	
	Block 1 General practice training		One day per week for:	Half-day per week for:
	Block 2 Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics		special training and reflection, in groups of trainees, at the training institute	self-directed learning
	Block 3 General practice training			
LEARNING OBJECTIVES	Block 1 Acquisition of knowledge, skills and attitudes with emphasis on common problems			
	Block 2 Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself			
	Block 3 Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations			

Since the introduction in 1973 of compulsory postgraduate training in general practice in The Netherlands, there has been discussion about the content, structure and duration of its curriculum.¹² The curriculum was extended from one year to two years in 1988 and again from two years to three years in 1994. With the start of the two-year curriculum the blueprint presented above has been introduced. The consequence of the extension from two years to three years was a change in the duration of the three blocks. In the two-year curriculum, each block took eight months. In the current three-year curriculum, each block takes one year. An argument for the extension of the curriculum was the assumption that the learning objectives of the training could not be achieved adequately in two years. A three-year curriculum should lead to a higher level of competence. This hypothesis is tested in our study, at least for knowledge. Research has shown that knowledge is core to any kind of clinical competence and that written test performance is predictive of performance in daily practice.^{9,13-19} We compared the

growth in knowledge of trainees in the two-year and three-year curricula and we tested whether there was a difference in level of knowledge between the curricula at completion of training. For this comparison, we have used longitudinally assembled assessment data of the two-year and the three-year training. Our assessment instrument was a written test that measures overall general practice knowledge relevant to daily practice.

Methods

Subjects

All trainees who received the regular two-year curriculum in the years 1992-1996 and the regular three-year curriculum in the years 1994-1999 were included in the study.

Instrument

The knowledge test of our study has a progress-testing format, so that the test intends to reflect the cognitive final objectives of the curriculum.^{20,21} These tests are administered periodically to all Dutch trainees in general practice at fixed intervals during their training to monitor progress. Naturally, a different test is used on every occasion. Each individual trainee will therefore be assessed a number of times and will demonstrate growth of knowledge as a function of training time. Since 1992, equivalent tests are constructed according to a blueprint, so that the content of each test contains a representative sample of the domain of general practice knowledge.⁹ The (multidimensional) blueprint is based on the Basic Job Description for the GP.¹¹ Items are classified according to the International Classification of Primary Care (ICPC)²² and four additional dimensions (age of patient, aspects of consultation, urgency of the problem and course of illness). Each test consists of 160 items. All items are embedded in vignettes representing general practice situations to increase their relevance to daily practice and assessment of crucial knowledge instead of knowledge of trivial facts. The response format is of the 'true-false-do not know' type. An example is shown in Box 2. A national test committee critically reviews the content and wording of all items. To discourage guessing, a score is calculated as the sum of the correct minus incorrect answers (formula scoring) and is expressed as a percentage of the maximum score. The reliability (Cronbachs' alpha) of the knowledge tests used in this study varied

from 0.61 to 0.76. The tests did not become more reliable over the years. These are satisfactory reliabilities, particularly for the interpretation of group results such as is used in this study.

Box 2

Example of two vignette-based knowledge items

Mr Van Bragten, aged 38 years, visits his GP. He is suffering from increasing deafness and buzzing in his right ear. The symptoms are not preceded by a common cold. Examination of the ear, nose and throat area does not reveal any abnormalities. Audiometry objectifies the deafness. The GP decides to refer Mr Van Bragten to the ENT specialist because she suspects an acoustic neurinoma.

The fact(s) in this case that match (es) with this diagnosis is/are:

- | | |
|--------------------------------------|-------|
| - the unilaterality of the symptoms; | T/F/? |
| - the buzzing in the ear. | T/F/? |

Reference: Huizing EH, Snow GB (red). Leerboek Keel-, Neus- en Oorheelkunde. Houten/Zaventem: Bohn Stafleu van Loghum; 1994: 152-3.

Procedure and analysis

A retrospective design was adopted. From 1992 to 1996, the test was administered to all Dutch trainees three times per year at fixed times. For practical reasons, the frequency has changed from three times to twice a year since 1996. Test results were used for formative purposes. For both curricula, we collected all available individual test scores and used these to calculate the average test score per training period. This mixed longitudinal design is considered an optimal design for the measurement of change over time.²³ As scores were sampled from different training cohorts, each average test score was composed of scores of different tests. In this way, potential differences in difficulty between tests were averaged. This was particularly important because no further actions were taken to standardise the test difficulties over the years. Variable starting times of the training programme and variable number of trainees per starting time resulted in a considerable variation in the number of trainees per training period and in the duration of training time per training period. To allow comparison

between training periods, the total training time was divided in equal periods of four months. Thus, training period one comprised test results of training months one-four, training period two comprised test results of training months five-eight, etc. Average test scores per training period were determined by assembling individual test scores within each training period. A range of four months was chosen for practical reasons. In the two-year curriculum, the average scores of six training periods were calculated (6 training periods \times 4 months = 24 months). In the three-year curriculum, the test was not administered in the ninth training period (9 training periods \times 4 months = 36 months). Therefore, the average scores of eight training periods were calculated. As a consequence, the level of knowledge at the very end of the three-year curriculum was not measured.

For the two-year and three-year curricula, the mean test scores and 95% confidence intervals were calculated per training period. Statistical difference was inferred from non-overlapping 95% confidence intervals.

Results

In total we collected 11952 test scores, obtained from 3682 different trainees (1563 of the two-year curriculum and 2119 of the three-year curriculum). The number of test scores per training period varied between 64 and 1849, and the number of tests used to calculate the average score per training period varied between 6 and 13.

The mean test scores with 95% confidence intervals of the two-year and three-year curricula are presented for each training period in Table 1 and Figure 1. For both curricula, the entrance level of knowledge was similar, and the greatest acquisition of knowledge occurred in the first eight months of training. This acquisition was significantly greater for the three-year curriculum. Thereafter, the level of knowledge in the two-year curriculum increased less rapidly but continuously, with a spurt towards the end of the curriculum. The growth increment was significant after each training period, except for periods four and five. The fourth training period covered the last four months of the second block, and in the fifth period trainees started their third block. In the three-year curriculum, the number of trainees in the third and sixth training period was rather small in comparison to the other training periods. Following a rapid increase, the acquisition of knowledge levelled off after training period two and increased rather slowly up to



Table 1

Mean correct minus incorrect scores as a percentage of the total and 95% confidence intervals (95%CI) per training period, based on data of trainees of the two-year and three-year curricula from 1992 to 1999

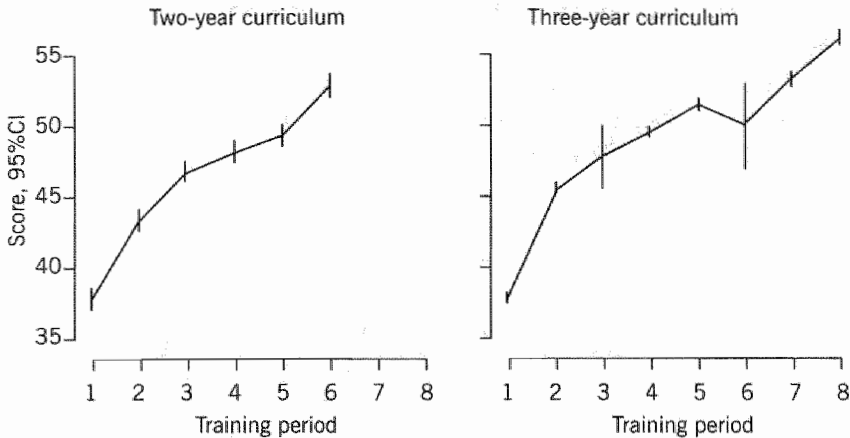
Training Period	Training time (months)	Two-year curriculum			Three-year curriculum		
		N	Correct-incorrect MEAN	95% CI	N	Correct-incorrect MEAN	95% CI
1	1-4	708	37.9	37.1 - 38.6	1849	37.6	37.2 - 38.0
2	5-8	593	43.4	42.5 - 44.2	1544	45.4	44.9 - 45.9
3	9-12	736	46.7	45.9 - 47.5	68	47.6	45.3 - 49.9
4	13-16	662	48.1	47.2 - 48.9	1394	49.3	48.8 - 49.8
5	17-20	689	49.2	48.4 - 50.0	1119	51.2	50.6 - 51.8
6	21-24	644	52.8	52.0 - 53.6	64	49.7	46.6 - 52.8
7	25-28				1057	52.9	52.3 - 53.5
8	29-32				824	55.8	55.1 - 56.5

and including training period seven to show again a rapid growth in training period eight. While the level of knowledge increased significantly between training periods two and five, in training period five and between training periods five and eight, this increase was rather small, especially in training periods six and seven. Training period six covered the last four months of the second block, and in the seventh training period trainees started their third block.

After one year of training (training period 3) and after two years of training (training period 6), both curricula showed comparable levels of knowledge. Finally, the level at the last measurement of the three-year curriculum (30 months training) was significantly higher than at the end of the two-year curriculum (24 months): 55.8% versus 52.8%.

Figure 1

Mean correct minus incorrect scores as a percentage of the total and 95% confidence intervals (95%CI) per training period and per curriculum



Discussion

This study tested the assumption that a three-year postgraduate training programme in general practice would lead to a higher competence level than a two-year programme. Competence was assessed by written progress-testing of general practice knowledge embedded in real life situations. Our results show that the overall level of knowledge with the three-year curriculum is higher after two-and-half years training than with the two-year curriculum after two years training. The difference (3%) is statistically significant and in our view meaningful when compared to a total increase in knowledge of 15% in the two-year curriculum and when compared to the average standard deviation of scores (10%). A three percent gain is then about one-third of the standard deviation of scores. The difference would probably have been larger if a real end measurement of the three-year curriculum had been available.

The study was conducted to provide insight not only into the level of knowledge at the end point of both curricula but also into the acquisition of knowledge during training. The results with the two-year curriculum show a rather continuous growth of knowledge with a spurt

at the beginning and at the end of the training programme and a levelling of at the end of the second block and the beginning of the third block. The pattern for the three-year curriculum is comparable to that of the two-year curriculum with respect to the greatest acquisition of knowledge at the beginning and towards the end of the training programme, and the rather continuous, but less rapid, increase during the rest of the training time. However, it is different in the duration of the period of smaller increase (20 months in the three-year curriculum versus 12 months in the two-year curriculum), and in the indication of a decline in knowledge in training period six. However, due to the low number of trainees in this training period, this finding has to be interpreted carefully.

In both curricula, most stagnation in growth occurs at the end of the second block, in which trainees rotate through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics, and at the start of the third block, where they return to training in general practice. The stagnation at the end of the second block can be explained by (1) a lack of acquisition of new general practice knowledge; (2) a lack of assimilation of this new knowledge due to a learning environment in which it is not needed; and (3) a loss of acquired general practice knowledge because there is no need for this knowledge to be used. All three explanations seem possible. It does not necessarily indicate that trainees do not acquire new knowledge and skills. Perhaps other knowledge, not measured by the test of this study, is acquired in this period or the acquisition of skills is more pronounced. Because the only insight we have into the contribution of the second block to the development of competence of trainees is the results of this study, further research into the second block is necessary. This research should aim to determine the nature and extent of its contribution to the development of competence of trainees and to the efficacy of the training as a whole. The stagnation in growth at the start of the third block is harder to explain. Perhaps trainees need time to become accustomed to their new learning environment or they make up for what they have missed.

When we compare the results of this study to those of a similar study into the growth of knowledge in medical school, we see that the level of knowledge at the start of postgraduate training approximates the level at the end of medical school (38% versus 40%).²⁴ At the end of postgraduate training, the level of knowledge has increased by 15 % in the two-year curriculum and by 18% in the three-year curriculum.

In our view, this a considerable increase in comparison to the increase of 36% found in undergraduate medical training across six years. To our knowledge, this is the first study into the growth of knowledge during postgraduate training in general practice. Although a controlled research design would give the best evidence,^{25,26} such a design is impossible when complete curricula are compared. In our study, we have aimed at the best available evidence by collecting over the years a large number of test scores for both curricula. Even with this design some methodological shortcomings remain. Firstly, the two curricula were in place at different times. Although learning objectives and the structure of the curricula remained the same, changes in content of the training programme seem inevitable due to new developments in the perception of good general practice and education. This drawback has been solved to some extent by the fact that each knowledge test is composed of recent knowledge supported by recent literature. Another time order effect could be due to changes in quality across trainee cohorts starting postgraduate training (e.g. changes in high school or in gender representation). This may influence learning styles and development of competence. As shown in our study, it did not influence the level of knowledge at the start of postgraduate training. Secondly, the two curricula would have been compared more adequately when the assessment of the trainees of the two-year curriculum was continued in their first year after completing their postgraduate training. By doing so, the effect of expiration of time without structured education on the development of knowledge could have been examined. Despite these shortcomings, this study has provided some insight into the effect of postgraduate training in general practice on the development of knowledge over time.

In conclusion, the extension of the two-year postgraduate training by one year, while maintaining the same structure and learning objectives, leads to a higher level of general practice knowledge. In the two-year and in the three-year curriculum, most acquisition of knowledge occurs in the first eight months of training and towards the end of training. Both curricula show the largest stagnation in growth at the end of the second block. This finding questions the impact of training in hospitals, clinics for chronically ill patients and psychiatric outpatients clinics on the development of competence of trainees in general practice and on the efficacy of the training as a whole. Further research into these aspects is needed before more firm conclusions can be drawn.

Acknowledgements

We would like to thank all trainees who took part in the study. We are grateful for the support from the staff of the eight Dutch institutes for postgraduate training in general practice. The members of the national committee of the knowledge test are thanked for their contribution to the knowledge tests. IJsbrand Kramer improved the English. This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

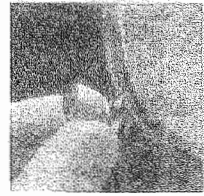


References

1. *Boerma WGW, De Jong FAJM, Mulder PH.* Health care and general practice across Europe. Utrecht: NIVEL; 1993.
2. *Grol R, Wensing M, Jacobs A, Baker R.* Quality assurance in general practice. The state of the art in Europe. Utrecht: NHG; 1993.
3. *Haq C, Ventres W, Hunt V, Mull D, Thompson R, Rivo M, Johnson P.* Where there is no family doctor: the development of family practice around the world. *Acad Med* 1995;70:370-80.
4. *Pabst R.* Medical education and reform initiatives in Germany. *Acad Med* 1995;70:1006-11.
5. *Campbell LM, Murray TS.* The effects of the introduction of a system of mandatory formative assessment for general practice trainees. *Med Educ* 1996;30:60-4.
6. *Zweifel J, Ringel M, Maudlin RK, Blossom HJ.* Extended educational sessions at three family residency programs. *Acad Med* 1996;71:1059-63.
7. *Cassie JM, Armbruster JS, Bowmer MI, Leach DC.* Accreditation of postgraduate medical education in the United States and Canada: a comparison of two systems. *Med Educ* 1999;33:493-8.
8. *Leigh TM, Young PR, Haley JV.* Performances of family practice diplomats on successive mandatory recertification examinations. *Acad Med* 1993;68:912-19.
9. *Van Leeuwen YD.* Growth in knowledge of trainees in general practice [dissertation]. Maastricht: Universitaire Pers Maastricht; 1995.
10. *Yudkowsky R.* Can resident evaluations demonstrate increase in residents' skills over time? *Acad Med* 1999;74(10Suppl):S108-10.
11. *Springer MP, editor.* Basic Job Description for the general practitioner. Utrecht: Dutch National Association of General Practitioners (LHV); 1983.
12. *Dubois V, Everwijn S, van Geldorp G et al.* The construction of a new curriculum of vocational training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
13. *Schmidt H, Norman G, Boshuizen H.* A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990;65:611-21.
14. *Bordage G.* Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med* 1994;69:883-5.
15. *Regehr G, Norman GR.* Issues in Cognitive Psychology: Implications for Professional Education. *Acad Med* 1996;7:988-1001.
16. *Van der Vleuten CPM.* The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
17. *Boshuizen HPA, van der Vleuten CPM, Schmidt HG, Machiels-Bongaerts M.* Measuring knowledge and clinical reasoning skills in a problem-based curriculum. *Med Educ* 1997;31:115-21.
18. *Ramsey PG.* Does clinical performance correlate with physicians' scores on written examinations? In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van de Steegh AFW, editors. *Advances in Medical Education*. Dordrecht: Kluwer Academic Publishers; 1997.

19. *Ram P, van der Vleuten C, Rethans J, Schouten B, Hobma S, Grol R.* Assessment in General Practice: the predictive value of written knowledge tests and a multiple station examination for actual medical performance in daily practice. *Med Educ* 1999;33:197-203.
20. *Kramer AWM, Pollemans MC.* Nationwide progress tests assessing knowledge in vocational training for general practice. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra PR, editors. *Teaching and Assessing Clinical Competence*. Groningen: BoekWerk Publ; 1990.
21. *Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW.* Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach* 1996;18:103-9.
22. *Lamberts H, Wood M, editors.* ICPC. *International Classification of Primary Care*. Oxford: Oxford University Press; 1987.
23. *Van 't Hof MA, Roede MJ, Kowalski CJ.* A mixed longitudinal data analysis model. *Hum Biol* 1977;49:165-79.
24. *Verhoeven BH, Verwijnen GM, Scherpbier AJJA, van der Vleuten CPM.* Growth of medical knowledge. *Med Educ* 2002;36:711-7.
25. *Harden RM, Grant J, Buckley G, Hart IR.* Guide No.1: Best Evidence Medical Education. *Med Teacher* 1999;21:553-62.
26. *Bligh J, Parsell G.* Research in medical education: finding its place. *Med Educ* 1999; 33:162-3.

Chapter 2



COMPARISON OF A RATIONAL AND AN EMPIRICAL STANDARD SETTING PROCEDURE FOR AN OSCE

This article has been published as:

Anneke Kramer, Arno Muijtjens, Koos Jansen, Herman Düsmen, Lisa Tan & Cees van der Vleuten. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Medical Education* 2003;37:132-139.

Summary

Purpose

Earlier studies of absolute standard setting procedures for objective structured clinical examinations (OSCEs) showed inconsistent results. This study compared a rational and an empirical standard setting procedure. Reliability and credibility were examined first. The impact of a reality check was then established.

Methods

The OSCE included 16 stations and was taken by trainees in their final year of postgraduate training in general practice and experienced general practitioners (GPs). A modified Angoff (independent judgements, no group discussion) with and without a reality check was used as a rational procedure. A method related to the Borderline Group procedure, the Borderline Regression (BR) method, was used as an empirical procedure. Reliability was assessed using generalisability theory. Credibility was assessed by comparing pass rates and by relating the passing scores to test difficulty.

Results

The passing scores were 73.4% for the Angoff procedure without reality check (Angoff I), 66.0% for the Angoff procedure with reality check (Angoff II) and 57.6% for the BR method. The reliabilities (expressed as Root Mean Square Errors) were 2.1% for Angoffs I and II, and 0.6% for the BR method. The pass rates of the trainees and GPs were 19% and 9% for Angoff I, 66% and 46% for Angoff II, and 95% and 80% for the BR method, respectively. The correlation between test difficulty and passing score was 0.69 for Angoff I, 0.88 for Angoff II and 0.86 for the BR method.

Conclusion

The BR method provides a more credible and reliable standard for an OSCE than a modified Angoff procedure. A reality check improves the credibility of the Angoff procedure but does not improve its reliability.

Introduction

In Objective Structured Clinical Examinations (OSCEs), performance is measured with checklists or rating scales. When the examination is used for decisive purposes, results must be compared to a standard of adequacy. Setting the standard is a judgmental process involving an arbitrary decision of what is considered to be 'good enough'.^{1,3} To guarantee that this process is defensible and controllable, several standard setting procedures have been developed.^{1,3} These procedures can be divided into two categories: relative (norm-referenced) and absolute (criterion-referenced) procedures.⁴ A relative standard is based on the performance of the test takers as a group (e.g. by establishing that the top 80% of examinees will pass). The absolute approach defines the passing score in terms of how many items or tasks have to be performed correctly to pass. Although relative procedures are easy to use and to explain, they have serious disadvantages. Firstly, the standard for a test will vary from group to group, depending on the ability of the group taking the test. Secondly, some examinees will pass and others will fail regardless of how correctly they perform and, as a consequence, the minimum standard of adequate performance will vary. As specific, well-defined tasks are measured in OSCEs, absolute standard setting procedures are more suitable, particularly when the results are to be used for certification.^{3,5} This study therefore focuses on absolute procedures.

Among absolute procedures, a fundamental distinction can be drawn between rational (test-centred) and empirical (examinee-centred) procedures.^{3,6} In a rational procedure, the standard is provided by expert judgement based on rational analysis of the test content. Examples of this approach are the Angoff and Ebel methods.¹ An advantage of the rational procedure is that it is founded on the content of the test; however, an accompanying risk is that standards may be set unrealistically high. Therefore, it is recommended that judges undergo a reality check by, for example, providing performance data.¹ In the empirical procedure, the standard is determined by judgement of the performance of individual candidates relative to a performance standard based on some external criterion or on overall test performance. Examples of this approach are the Contrasting Groups and Borderline Group methods.¹ The empirical procedure provides a more holistic approach to standard setting and seems therefore particularly appropriate for performance tests in which a few, relatively

lengthy tasks are assessed.⁶

Both procedures have been subject to study and evidence suggests a considerable variability in standards when different methods and judges are used.³ There is also evidence that reliability varies among different methods.³ Studies examining the impact of performance data on standards for written tests found contradictory results. Some results indicated that performance data reduced the range of judgements, while others did not find it to have any influence.^{4,6} Several studies have compared rational and empirical standard setting procedures for OSCEs. Some studies demonstrate higher pass rates for the empirical procedure,^{3,7,8} while others show higher pass rates for the rational procedure.^{3,9,10} Kaufman et al. found comparable passing scores for an Angoff and a Borderline Group method.¹¹ They also investigated the reliability of the Angoff method. Their conclusion was that a defensible and feasible passing score could be established using both methods, but that a large number of judges or stations would be required to obtain acceptable reliability for the Angoff procedure.

The need for more research into rational and empirical standard setting procedures for OSCEs is obvious. In particular, issues concerning the impact of a reality check and reliability require further investigation. Therefore this study focused on the following research questions:

1. What levels of reliability are shown by rational and empirical standard setting procedures in an OSCE and what is the impact of a reality check on the reliability of the rational procedure?
2. What is the credibility of both procedures and what is the effect of a reality check on the credibility of the rational procedure?

We chose to use a modified Angoff method for the rational procedure. It was modified so that, unlike the typical Angoff procedure, no group discussion followed by a second judgement procedure was applied. The empirical procedure consisted of a method related to the Borderline Group method and will be referred to as Borderline Regression (BR) method. In the BR method OSCE examiners rate clinical performance on a global rating scale. Checklist scores are subsequently regressed on the global ratings. The resulting equation is then used to calculate the checklist passing score. Woehr et al. described a similar procedure using the relationship between test scores and criterion performance data to arrive at a passing score.¹² The advantage of the BR method is that it uses the complete rating scale rather than the classes of performance used by the Borderline

Group or Contrasting Groups methods. Reliability of the Angoff method was assessed using generalisability theory. A new procedure was developed to examine the reliability of the BR method. The procedure was based on a division of the available data into a number of random subsamples that were treated as representations of a group facet in a generalisability study. Credibility was judged by comparing the pass/fail rates of the different methods for a group of trainees in general practice and a reference group of experienced general practitioners (GPs). We considered that if a passing score was to be credible, a majority of GPs should pass (e.g. 70%). Because test procedures, including OSCEs, differ in levels of difficulty, we also related station passing scores derived from both methods to station difficulty. In our opinion, a procedure that is sensitive to station difficulty is more credible.

Methods

The OSCE

The OSCE was taken by 86 trainees in their final (third) year of training, randomly selected from the eight Dutch postgraduate training institutes in general practice. In addition, 35 GPs who also serve as trainers volunteered to act as the reference group for the examination.

The OSCE included 16 stations, representing a cross-section of the domain of clinical skills in general practice. Eight stations involved manikins, while eight involved standardised patients. Seven-minute stations were scheduled for seeing a patient with impaired vision, for insertion of a naso-gastric tube, for resuscitation, for urinary catheterisation, for seeing a patient with painful micturition, for intravenous cannulation, for injection into the shoulder and for laboratory investigation of vaginal discharge. Fifteen minute stations were scheduled for seeing a patient with a painful knee, a patient with a fish bone in the throat and a patient with backache, for testing pulmonary reversibility for diagnosing asthma, for physical examination of the female breast, for wound suturing with anaesthesia, for applying a compression bandage for a venous leg ulcer and for annual control of a patient with diabetes mellitus type 2.

Eighty-four examiners were involved, all of whom were experienced GPs as well as teachers at the training institutes. Most of them were familiar with the OSCE format. Each examiner was trained in the rating of two different stations. The same examiners rated the same two stations. The purpose of the training was to reach a consensus in



scoring between raters. A task-specific checklist, developed from national guidelines for general practice, was used to score clinical performance.^{13,14} In addition to the checklist, performance was assessed using a 10-point global rating scale.^{8,10} The examiners were made aware that a global rating below 5.5 represented inadequate performance. The checklist score of a station was defined by the percentage of correctly performed items on the checklist. The total test score was calculated by averaging the 16 station scores. The global rating was only used for standard setting purposes. Using the checklist score, the norm oriented reliability of the test was 0.66 (generalisability coefficient) and the domain oriented reliability was 0.59 (dependability coefficient).

The judges

The OSCE examiners served as judges for the standard setting procedures. They participated on a voluntary basis.

The rational standard setting procedure

A modified Angoff procedure was applied to estimate the passing score on item content.¹ In order to include a reality check, the procedure was performed twice: once before the rating training (Angoff I) and once directly after the examination (Angoff II). Judgements were made individually using detailed written instructions. For purposes of efficiency and feasibility, no opportunity for discussion and an adjusted second judgement were included in the procedure. The judges were asked to estimate for each item on the checklist which proportion of the borderline candidates would correctly perform this item. The borderline candidate was defined as a candidate who performs at a level between pass and fail. The Angoff passing score was calculated for each station by averaging the estimations across judges and items. The Angoff passing score of the total test was defined by averaging the 16 station passing scores.

The empirical standard setting procedure

The BR method was applied to establish a passing score based on the empirical approach. By using the global ratings of the overall performance (on the 10-point scale), the pass/fail borderline was defined at 5.5 on the scale (5.5 is the traditional borderline for the Dutch marking system). The corresponding passing score per station on the checklist score scale was obtained by regressing the checklist

scores on the global ratings, and then calculating the checklist score on the regression line for the global rating set at 5.5. The corresponding BR passing score for the total test was defined by averaging the 16 station passing scores.

Statistical analysis

Descriptive analysis

Mean OSCE checklist and global rating scores and standard deviations of trainees and GPs were calculated.

Mean Angoff I, Angoff II and BR passing scores were calculated. The results were tested for statistically significant differences using a paired t-test.

Reliability of the rational standard setting procedure

The reliability of the Angoff procedure was assessed using generalisability theory.¹⁵ The test, consisting of 16 stations, was divided into eight pairs of stations and each pair was judged by a different subset of raters (judges nested within station pairs). All raters in the subset judged each of the two stations (judges were crossed with stations within a couple of stations). A repeated generalisability analysis was performed to allow for a crossed design analysis. A crossed station-by-judge ANOVA design ($S \times J$) was used for each pair of stations and variance components were estimated using the GENOVA package.¹⁶ An overall estimate for each variance component was obtained by averaging the corresponding variance estimations across the station pairs.

Because we wished to estimate the error of setting a standard for a given test, the variance of the station main effect was not included in the passing score error variance. The Root Mean Squared Error (RMSE) of the Angoff passing score was estimated accordingly by

$$RMSE_{ANG}^2 = \sigma_J^2 / n_J + \sigma_{SJ}^2 / (2 \times n_J) \quad (1)$$

where n_J is the number of judges and σ is the estimated variance component for the associated effect (J : judge main effect; SJ : station-judge interaction effect). $RMSE_{ANG}$ is an estimate of the standard error of the average Angoff passing score across judges and stations, assuming equal numbers of judges per pair of stations. Because judges were nested within station pairs, the number of station-judge

combinations in the design equals $2 \times n_j$ instead of $n_s \times n_j$ as with a crossed design. This implies that when the number of judges is not changed, an increase of the number of stations in the test does not result in a change in the predicted precision of the passing score (i.e. the number of stations has no direct influence on the precision of the passing score). Using the variance components σ_j and σ_{sj} obtained in the generalisability analysis, the error in the test's Angoff passing score was calculated according to equation (1) for several numbers of judges ($n_j = 16, 48, 80, 160$). These numbers of judges correspond to nested designs with 1, 4, 10, or 20 raters per station pair, respectively, in a 16 station test.

Reliability of the empirical standard setting procedure

The reliability of the BR method was also investigated. A newly developed procedure was used in order to obtain information about the accuracy of the BR passing score estimate. In this procedure, the sample of 86 trainees was split up into random subsamples of, for instance, eight groups of 10 trainees per group. The BR passing score per station for each subsample was calculated as described before. The resulting 16×8 (station \times subsample) matrix of BR passing scores was used as input for a station-by-subsample ANOVA analysis in the GENOVA package.¹⁶ The analysis resulted in estimates of the variance components associated with stations (σ_s^2), subsamples (σ_G^2) and their interaction (σ_{sG}^2). The passing score of the test was obtained by averaging the passing scores of the stations. Accordingly, the RMSE of the passing score ($RMSE_{BR}$) was defined by

$$RMSE_{BR}^2 = \sigma_G^2 + \sigma_{sG}^2 / n_s \quad (2)$$

with n_s representing the number of stations in the test. Note that $RMSE_{BR}$ represents the estimation error of the BR passing score when the estimation is based on the data of a single subsample.

The random split up of the original sample of 86 trainees was repeated for varying subsample sizes (seven subsamples of size 12, six of 13, five of 17, four of 20 and three of 29). For each subsample size the variance components σ_G^2 and σ_{sG}^2 were estimated and the corresponding $RMSE_{BR}$ was calculated according to equation (2) for varying numbers of stations ($n_s = 1, 4, 8, 12, 16, 20$).

Thus, for each number of stations a series of six $RMSE_{BR}$ was obtained corresponding to the six subsample sizes used (10, 12, 13, 17, 20 and 29). Generally, the error variance of a parameter estimate is inversely proportional to the sample size. Using this relation, the $RMSE_{BR}$ for the original sample size (86 trainees) was calculated by extrapolating the series of six $RMSE_{BR}$ values for each number of stations.¹⁷ The same relationship was used to calculate the $RMSE_{BR}$ for several other (hypothetical) numbers of trainees (40, 160 and 320).

Credibility

Pass/fail rates of Angoff I, Angoff II and BR passing scores were calculated for trainees and GPs.

Pearson correlations of the station passing scores and p-values (mean percentage correctly performed items per station) were calculated for Angoff I, Angoff II and BR method.

Results

The OSCE scores of the trainees were higher than those of the GPs for both the checklist and global rating scores (68.6% versus 63.3% and 7.0 versus 6.8, respectively) (see Table 1).

The mean Angoff I passing score of the total test was 73.4%. The mean Angoff II passing score was found to be significantly lower: 66.3% ($n=16$, $p<0.001$). The mean BR passing score of the total test was 57.6%. This passing score was significantly lower than both Angoff passing scores ($n=16$, $p<0.001$). Table 2 shows the mean estimated variance components for Angoff I and II.

Table 1
Mean OSCE checklist and global rating scores and standard deviations of trainees and GPs

Examinees	N	OSCE Checklist score	OSCE Global rating score
		MEAN (SD)	MEAN (SD)
Trainees	86	68.6 (6.1)	7.0 (0.49)
GPs	35	63.3 (7.0)	6.8 (0.5)

Table 2

Estimated variance components of the Angoff I and Angoff II passing scores

Source	Angoff I			Angoff II		
	Estimated variance component	Standard error of estimated variance component	Percentage of total variance	Estimated variance component	Standard error of estimated variance component	Percentage of total variance
Judges (J)	0.37	14.6	0.03	0.67	15.2	0.06
Stations (S)	815.2	303.4	66.4	775.9	290.8	64.4
SJ,e	411.7	66.4	33.5	427.2	68.9	35.5

Due to practical reasons, only 51 of the 84 examiners served as judges in both Angoff procedures, with a minimum of four and a maximum of 11 judges per pair of stations. The generalisability analysis for the Angoff procedure was based on the 51 examiners that were involved in Angoff I as well as Angoff II.

For both Angoff I and II passing scores, approximately two thirds of the total variance was attributable to variation among stations, indicating a wide range of difficulty levels across stations. This implies that the universal Angoff passing score (i.e. the true Angoff passing score) may vary considerably for tests comprising small numbers of stations. The percentage of variance associated with judges was very small, while the variance component corresponding to station-judge interaction accounted for approximately one third of the total variance. This indicates that the judge main effect in the Angoff passing scores is minor.

Table 3 presents the predicted Root Mean Squared Errors (RMSE) of the test's passing score of Angoff I and II as a function of the number of judges in the test. The RMSE is the error of the test's passing score expressed on the scoring scale. With 48 judges (approximately the number of judges that participated in both Angoff procedures), an RMSE of 2.1% on the scoring scale was achieved for Angoff I and II. Multiplying the RMSE by 1.96 yields a 95% confidence interval for the test's passing score. For the passing score found in our OSCE,

Table 3
Root Mean Squared Error of Angoff I and II passing scores as a function of the number of judges

Angoff	Number of judges			
	16	48	80	160
Angoff I	3.6	2.1	1.6	1.1
Angoff II	3.7	2.1	1.6	1.2

this corresponds to an interval of $73.4\% \pm 4.1\%$ for Angoff I and an interval of $66.0\% \pm 4.1\%$ for Angoff II. This results in pass rates for trainees ranging from 9.3% to 45.3% for Angoff I and from 41.9% to 83.7% for Angoff II.

Table 4 shows the RMSEs of the test's passing score of the BR as a function of the number of trainees and stations in the test. With 16 stations and 86 trainees, an RMSE of 0.6 was achieved. This yields a confidence interval of $57.6\% \pm 1.2\%$, resulting in pass rates for trainees ranging from 91.9% to 97.7%.



Table 4
Root Mean Squared Error of the BR passing scores as a function of the number of trainees and stations

Number of trainees	Number of stations					
	1	4	8	12	16	20
40	3.3	1.7	1.2	1.0	0.9	0.8
86	2.3	1.2	0.8	0.7	0.6	0.6
160	1.7	0.9	0.6	0.5	0.5	0.4
320	1.2	0.6	0.4	0.4	0.3	0.3

Table 5
Pass rates of trainees and GPs for the Angoff I, Angoff II and BR passing scores

Procedure	Passing score %	Pass rate	
		Trainees (N=86) %	GP (N=35) %
Angoff I	73.4	18.6	8.6
Angoff II	66.3	66.3	45.7
BR	57.6	95.3	80

Table 5 presents the percentages of trainees and GPs passing the examination when each of the three passing scores is applied. The highest pass rate was obtained by the BR method. The lowest pass rate was obtained by Angoff I, only 9% of the GPs and 19% of the trainees would have passed.

The correlation of the station difficulties (p-values) and the station passing scores was 0.69 for Angoff I, 0.88 for Angoff II and 0.86 for the BR method.

Discussion

This study concerns issues of reliability and credibility for a modified Angoff procedure and an empirical procedure related to the Borderline Group method, the Borderline Regression (BR) method.

Angoff procedure

This procedure, which is based on the content of the test and the concept of the hypothetical borderline candidate, has a tendency to set the standard too high.^{1,4,18} We therefore applied the procedure *with* and *without* a reality check (respectively, Angoff II and Angoff I). To obtain a feasible procedure, we had to deviate from the typical Angoff procedure by omitting the group discussion and adjusted second judgement part of the procedure. The reality check proved to have

hardly any influence on the reliability. For both Angoff I and Angoff II, two thirds of the estimated variance was attributable to station variance, while the estimated judge variance was very low ($< 0.1\%$). The estimated overall variance (including the interaction effect between judges and stations) was one third. In the estimation of the error involved in the test's passing score, the station variance component was not considered part of the error, because the procedure is aimed at estimating the passing score associated with a certain set of stations. With 80 judges (approximately the number of examiners in the test) the error in the passing score would still amount to 1.6% on the scoring scale. This implies a 95% confidence interval of a width equal to $2 \times 3.2\%$. Given that with normally distributed test scores a 1% shift in the passing score changes the failure rate by approximately 2.5%, this error is too high. Using more judges would reduce the error but from a resource perspective this hardly seems feasible in regular practice.

Credibility was assessed by comparing pass rates. Only 19% of the trainees and 9% of the experienced GPs would pass with Angoff I. We do not consider such high failure rates to be credible for a decisive examination and for experienced GPs. The reality check improved credibility: 66% of the trainees and 46% of the GPs would have passed. However, the failure rates are still rather high. When credibility was examined by investigating to what extent the procedure takes test difficulty into account, the station passing scores of Angoff II had a higher correlation (0.88) with the station checklist scores than did those of Angoff I (0.69). This indicates that Angoff II is more sensitive to test difficulty than Angoff I. This should not be surprising as the reality check passes station difficulty information to the judges. Given the high correlation of passing scores and station difficulty for both Angoff I and Angoff II, it is also not surprising to find a large station component in the variance component estimation (Table 2). This seems to be a natural consequence of variability in station difficulty, as is usually found in OSCEs. Apparently, this is appropriately reflected in the passing scores. From this perspective, the large station variance component in the Angoff passing score seems a desirable outcome.



Borderline Regression procedure

This procedure, based on the overall judgement of the actual performance by the examiners, provides a more lenient and holistic concept of competence than the Angoff procedure.^{6,8}

A reality check is inherent to the method. The current study confirms this leniency because pass rates would have been 95% for trainees and 80% for GPs. This finding supports the credibility of the procedure as a vast majority of passes is expected for GPs as well as for trainees at the end of their postgraduate training. Credibility was also supported by the high correlation of the station passing scores with the station checklist scores (0.86).

The error in the test's passing score has an acceptable precision level (95% confidence interval of a width equal to $2 \times 1.2\%$) with the number of stations and trainees used in the test, implying a pass rate ranging from 92% to 98%.

Methodological considerations

Our application of the Angoff procedure, without group discussion and adjusted judgement, may lead to less precise and more variable estimates, resulting in a lower level of reliability than the typical Angoff. However, important disadvantages of the typical application are that it is organisationally complex, time-consuming and expensive, particularly when the procedure has to be applied twice for one test, as would have been required in our study.

Different designs were applied when estimating the error involved in the Angoff and BR passing scores. This difference is a consequence of the nature of both procedures. In the Angoff procedure the passing score is obtained by judging the content of the stations, while in the BR procedure the passing score is estimated by judging the performance level of the trainees. Consequently, the size of the passing score error in the Angoff method is determined by the number of judges, whereas in the BR method, it is determined by the number of stations and trainees. Nevertheless, both RMSEs indicate the size of the passing score error when the two standard setting methods are applied in a practically feasible manner. Hence, the results help to decide which of the two procedures is preferable, given the restricted availability of resources in regular practice.

Generalisation of the results obtained in the current study is limited, because the study was based on a single test and the numbers of trainees and judges were rather small.

Comparison of the Angoff and Borderline Regression procedures

On the basis of the results in this study, the BR procedure seems a more credible and reliable procedure with which to set a standard for

an OSCE than does the modified Angoff procedure, even when a reality check is applied to the latter.

Our findings support discussion in the literature that the Angoff procedure may be less convenient for performance-based skills testing than for written knowledge testing. In the Angoff method the proportion correct of borderline examinees has to be estimated for each item, assuming that each item is content-independent. However, questions remain as to what extent this can be true for the OSCE.^{6,19,20}

The BR method focuses on overall performance per station rather than performance per item.

From a cost perspective, the BR method is preferable, although our application of the Angoff method (written instruction only, individual estimation without discussion) represents a good alternative.

Conclusion

We conclude that the BR method provides a reasonable and defensible standard for an OSCE and that the method is practically feasible.

Acknowledgements

We like to thank all trainees and GPs who took part in the study. We are grateful for the support from the staff of the eight Dutch institutes for postgraduate training in general practice and to the members of the national test committee of the OSCE for their contribution to the development of the test. Thanks are also due to IJsbrand Kramer for polishing our English. This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).



References

1. *Livingstone SA, Zieky MJ.* Passing Scores. Princetown: Educational Testing Service; 1982.
2. *Norcini J.* Approaches to Standard-Setting for Performance-Based Examinations. In: Harden RM, Hart IR, Mulholland H, editors. Approaches to the Assessment of Clinical Competence Part I. Proceedings of the 5th Ottawa International Conference on Medical Education and Assessment; 1992 Sept 1-3; Dundee, Scotland. Dundee: Centre for Medical Education; 1992:32-7.
3. *Cusimano M.* Standard setting in medical education. Acad Med 1996; 71(10Suppl):S112-20.
4. *Norcini JJ.* Research on standards for professional licensure and certifications examinations. Eval Health Prof 1994;17:160-77.
5. *Newble D, Jolly B, Wakeford RE, editors.* The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence. Cambridge: University Press; 1994.
6. *Kane MT, Crooks TJ, Cohen AS.* Designing and Evaluating Standard Setting Procedures for Licensure and Certification Tests. Adv Health Sci Educ 1994;4:195-207.
7. *Van Luijk SJ, van der Vleuten CPM.* A Comparison of Standard Setting Methods Applied to a Performance-based Test. In: Harden RM, Hart IR, Mulholland H, editors. Approaches to the Assessment of Clinical Competence Part I. Proceedings of the 5th Ottawa International Conference on Medical Education and Assessment; 1992 Sept 1-3; Dundee, Scotland. Dundee: Centre for Medical Education; 1992:326-30.
8. *Rothman AI, Cohen R.* A comparison of empirically- and rationally-defined standards for clinical skills checklists. Acad Med 1996;71(10Suppl):S1-3.
9. *Norcini JJ, Stillman PL, Sutnick AI et al.* Scoring and standard setting with standardized patients. Eval Health Prof 1993;16:322-32.
10. *Dauphinee WD, Blackmore DE, Smee S, Rothman AI, Reznick R.* Using the judgments of physician examiners in setting the standard for a national multi-center high stakes OSCE. Adv Health Sci Educ 1997;2:201-11.
11. *Kaufman DM, Mann KV, Muijtens AMM, van der Vleuten CPM.* A Comparison of standard-setting procedures for an OSCE in undergraduate medical education. Acad Med 2000;75:267-71.
12. *Woehr DJ, Arthur W, Fehrmann ML.* An empirical comparison of cutoff score methods for content-related and criterion-related validity setting. Educ Psy Meas 1991;51:1029-39.
13. *Grol RPTM.* National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. Br J General Pr 1990;40:361-4.
14. *Jansen JJM, Tan LHC, van der Vleuten CPM, van Luyk SJ, Rethans JJ, Grol RPTM.* Assessment of competence in technical clinical skills of general practitioners. Med Educ 1995;29:247-53.
15. *Brennan RL.* Elements of Generalizability Theory. Iowa City: ACT Publications; 1983.

16. *Crick JE, Brennan RL.* A general purpose analysis of variance system [computer program]. Version 2.2. Iowa City: The American College Testing Program; 1984.
17. *Muijtjens AMM, Kramer AWM, Kaufman DM, van der Vleuten CPM.* Using resampling to estimate the precision of an empirical standard setting method. Accepted by Appl Meas Educ 2003.
18. *Norcini J.* The Credibility and Comparability of Standards. Appl Meas Educ 1997;10:3959.
19. *Clauser BE, Clyman GC.* A Contrasting-Groups Approach to Standard Setting for Performance Assessments of Clinical Skills. Acad Med 1994;69(10Suppl):S42-4.
20. *Ross LP, Clauser BE, Margolis MJ, Orr NA, Klass DJ.* Standard Setting. An expert-judgement approach to setting standards for a standardized-patient examination. Acad Med 1996;71(10Suppl):S4-6.



100

The first part of the chapter discusses the importance of the

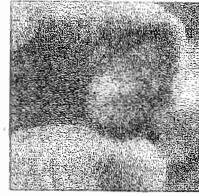
second part of the chapter discusses the importance of the

third part of the chapter discusses the importance of the

fourth part of the chapter discusses the importance of the

fifth part of the chapter discusses the importance of the

Chapter 3



PREDICTIVE VALIDITY OF A WRITTEN KNOWLEDGE TEST OF SKILLS FOR AN OSCE IN POSTGRADUATE TRAINING FOR GENERAL PRACTICE

This article has been published as:

AWM Kramer, JJM Jansen, P Zuithoff, H Düsman, LHC Tan, RPTM Grol & CPM van der Vleuten. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. *Medical Education* 2002;36:812-819.

Summary

Purpose

To examine the validity of a written knowledge test of skills for performance on an OSCE in postgraduate training for general practice.

Methods

A randomly selected sample of 47 trainees in general practice took a knowledge test of skills, a general knowledge test and an OSCE. The OSCE included technical stations and stations including complete patient encounters. Each station was checklist rated and global rated.

Results

The knowledge test of skills was better correlated to the OSCE than the general knowledge test. Technical stations were better correlated to the knowledge test of skills than stations including complete patient encounters. For the technical stations the rating system had no influence on the correlation. For the stations including complete patient encounters the checklist rating correlated better to the knowledge test of skills than the global rating.

Conclusion

The results of this study support the predictive validity of the knowledge test of skills. In postgraduate training for general practice a written knowledge test of skills can be used as an instrument to estimate the level of clinical skills, especially for group evaluation, such as in studies examining the efficacy of a training programme or as a screening instrument for deciding about courses to be offered. This estimation is more accurate when the content of the test matches the skills under study. However, written testing of skills cannot replace direct observation of performance of skills.

Introduction

Clinical skills constitute an important part of the daily work of general practitioners (GPs).^{1,2} Undergraduate medical training and postgraduate training in general practice are intended to prepare trainees adequately for this task. However, the efficacy of medical training for the acquisition of clinical skills has not been extensively evaluated. There are indications that the quality of skills training is not always satisfactory.^{3,6} Assessment of clinical skills in postgraduate training for general practice is therefore of great importance to evaluate the training programme for mastery of clinical skills.

Various methods have been developed for assessment of clinical skills. Among them the Objective Structured Clinical Examination (OSCE) has been identified as one of the most valid methods for the assessment of clinical skills both in undergraduate and postgraduate training.

However, it has a serious drawback.⁷⁻¹¹ Research has shown a large variability in performance between cases or tasks, indicating a high content-specificity of the skills assessed.¹² Consequently, many stations are needed to reach reliable test results. The organisation of an OSCE is complex and expensive, especially when many stations are included.¹³

Due to these logistical limitations, large-scale application of this method is restricted. In recent years, several studies have therefore focused on alternative instruments with less demanding characteristics.^{2,3,14-17} In these studies the relationship between a written knowledge test of skills, a self-assessment questionnaire and a GP-trainer's report, and an OSCE have been investigated. The results of these studies are summarised in Table 1.

These studies show the existence of a reasonably high true correlation between a written knowledge test of skills and an OSCE, whereas a moderate correlation is found for a self-assessment questionnaire. A GP-trainer's assessment of skills does not predict at all performance on an OSCE. From these results it can be concluded that a written knowledge test of skills seems to have a reasonable predictive value for performance on an OSCE in medical school and in continuing medical education in general practice. The question remains whether this is true for postgraduate training in general practice. Therefore, this study focused on the predictive validity of a written knowledge test of skills in postgraduate training for general practice. We restricted the study to the assessment of clinical skills (history taking, physical examination, diagnosing and management); communicative and

interpersonal skills were not included. Three hypotheses were tested. Our first hypothesis concerned the specific content of a knowledge test and how this determines its predictive value for an OSCE. In a knowledge test of skills only those aspects of knowledge are assessed which have a clear relationship with the conduct of clinical skills. Therefore, it was hypothesised that a knowledge test with this content would be better correlated to an OSCE than a general knowledge test that covers a broader cognitive medical domain (*hypothesis 1*). Our second hypothesis concerned the extent to which the rating system of the OSCE determines the correlation with a knowledge test of skills. Originally, detailed checklists were used to rate performance and although these provide valid and reliable results, they are not without risks.¹⁸ Examples are the findings that excessive detail of the checklists no longer reflects clinical reality. Moreover, it conditions the students to memorise isolated items and perform accordingly, obliterating the need to deepen the understanding of the skills.¹⁹ Finally, by the fragmentation of performance in isolated items some aspects are not assessed, e.g. the student's attitude to the patient and the logical sequencing and suppleness of the consultation. Aspects that are particularly important at the postgraduate level. These aspects are best examined by a professional judgement of experienced doctors on holistic attributes in a rating scale.²⁰ To circumvent this problem, global scales rated by experienced doctors were introduced as an alternative to the detailed checklists and these are increasingly and successfully used in OSCEs, in combination or not with checklists.^{10,18-20} It was hypothesised that a knowledge test of skills would be better correlated to an OSCE rated with checklists than to an OSCE rated with global scales (*hypothesis 2*). A checklist corresponds better to a knowledge test of skills, because its content is more detailed and task-specific than a global rating scale. Our last hypothesis concerned the content of the OSCE and the knowledge test of skills. In OSCEs several skill aspects can be assessed, varying from a very specific technical task (e.g. writing out a prescription) to a complete patient encounter, including history taking, physical examination and treatment. The knowledge test of skills of our study included the more technical, procedural aspects of clinical skills.²¹ It was therefore hypothesised that it would be better correlated to stations including technical, procedural skills than to stations including complete patient encounters (*hypothesis 3*).

Table 1
Summary of the studies discussed

Reference	N	Educational level	Alternative instrument	OSCE	Observed correlation	True correlation
Verhoeven et al. ¹⁶	38	Third year Problem-based medical school	Knowledge of skills True/false/? 86 items	Checklist 2 hours 8 stations blueprint unknown	0.65	1.0
Scherpbier ³	59	Fourth and final (6th) year Traditional and problem-based medical school	Knowledge of skills True/false/? 111 items	Checklist Global rating 1.5 hours different blueprint	0.47 0.35	0.86 0.65
Van der Vleuten et al. ¹⁴	27	Final year Problem-based medical school	Knowledge of skills True/false/? 238 items	Checklist 2 hours 8 stations same blueprint	0.72	0.89
Remmen et al. ¹⁷	47	Final year Traditional medical school	Knowledge of skills True/false/? 132 items	Checklist Global rating 2.5 hours 12 stations same blueprint	0.48 0.41	0.87 0.70
Remmen et al. ¹⁷	59	Final year Traditional medical school	Knowledge of skills True/false/? 132 items	Checklist Global rating 2.5 hours 12 stations same blueprint	0.41 0.35	0.85 0.70
Jansen et al. ¹⁵	96	Experienced GPs	Knowledge of skills True/false/? 75 items	Checklist 2 hours 8 stations same blueprint	0.33	0.77
Jansen et al. ¹⁵	96	Experienced GPs	Self-assessment questionnaire 20 items	Checklist 2 hours 8 stations same blueprint	0.29	0.47
Ram et al. (personal communication)	90	Experienced GPs	Knowledge of skills True/false/? 124 items	Maas-global 1.5 hours 8 stations different blueprint	0.35	0.57
Kelly et al. ²	25	Completing postgraduate training for GP	GP-trainer's report	Checklist 8 stations same blueprint	A significant number of GP-trainees were unable to perform correctly on the OSCE despite receiving a satisfactory trainer's report	

We tested the three hypotheses by examining the Pearson correlation between a knowledge test of skills and a general knowledge test, respectively, and an OSCE. The OSCE included technical and complete patient encounter stations.

Methods

Subjects

From a total number of 237 GP-trainees in their final (third) year of postgraduate training, a randomly selected sample of 47 trainees took a knowledge test of skills, a general knowledge test and an OSCE. The sample did not significantly differ from the rest of the trainees on the aspects age, time between graduation and start postgraduate training, entree level of general knowledge and knowledge of skills (t-test, $p > 0.05$). The percentage of male trainees in the sample was somewhat higher (49% vs. 42%). All eight Dutch postgraduate training institutes were represented with a minimum of three and a maximum of nine trainees.

Instruments

Three different instruments were used to assess competence in skills: a knowledge test of skills, a general knowledge test and an OSCE. National test committees developed the tests, using national guidelines for general practice and consensus procedures.²² Members of the test committees were experienced GPs, with extensive teaching experience in postgraduate training, from the eight training institutes and test experts. All tests reflected the final objectives of the curriculum. The *knowledge test of skills* (KTS) contained 180 items in the 'true/false/do not know' response format. The test was constructed according to a blueprint based on the International Classification for Primary Care (ICPC)²³ and a list of clinical skills relevant for general practice,²⁴ implying that the content of the test represented a cross-section of the domain of clinical skills relevant to general practice. Knowledge of skills was defined as knowledge concerning the technical, procedural aspects of skills and knowledge assessing skill-related aspects (e.g. materials and indications).²¹ The items were, as far as possible, embedded in vignettes representing general practice situations to increase relevancy for daily practice and assessment of crucial knowledge rather than knowledge of trivial facts. In Box 1 an example is presented.

Box 1

Example of a vignette-based knowledge of skills item

Marco Terborg, aged 5 years, visits his GP with his father. Marco's father is afraid that Marco's right leg is longer than his left leg. The GP decides to objectively measure the length of Marco's legs. For this purpose she measures on both sides the distance between the spina iliaca anterior superior and the lateral malleolus.

- This is a correct method to measure the leg length. T/?/F

Reference: Visser JD. Kinderorthopaedie: pluis of niet pluis. Groningen: Van Denderen; 1996:55-61

The *general knowledge test* (GKT) consisted of 160 items. The item format was similar to the KTS, only the content differed (see Box 2). The test was constructed according to a blueprint based on the Basic Job Description for the GP, implying that the content of the test represented a cross-section of the domain of overall general practice knowledge.^{25,26}

Box 2

Example of two vignette-based general knowledge items

GP Dr de Boer visits Mr Van Loen, aged 45 years, at home. For two hours Mr Van Loen has been suffering from an acute retrosternal pain. For three days prior to this he has been ill with symptoms indicating influenza. He has no significant past history. Physical examination reveals that he is breathing shallowly and that the pain increases on inspiration and decreases when sitting forwards. His temperature is 38.6° C. His blood pressure and pulse rate are normal. Auscultation of the lungs and heart is normal.

The GP wishes to differentiate between a pericarditis and a myocard infarction. The following symptom(s) is/are more in favour for a pericarditis than for a myocard infarction:

- the increase of pain at inspiration; T/?/F
- the decrease of pain when sitting face forward. T/?/F

Reference: Roeland JRTC, Lie KI, Wellens HJJ, Van de Werf F. Leerboek cardiologie. Houten: Bohn Stafleu Van Loghum; 1995:418-9.

For both tests the score was calculated as the sum of the correct minus incorrect answers to discourage guessing (formula scoring) and was expressed as percentage of the maximum score. The reliability (Cronbachs' alpha) was 0.76 for the KTS and 0.70 for the GKT. The total OSCE included 16 stations, representing a cross-section of the clinical skills relevant for general practice spread among 16 ICPC-categories.^{23,24} The total testing time was three hours. Nine stations involved specific technical, procedural tasks (e.g. intravenous injection), while in seven stations complete patient encounters were assessed (e.g. a patient with backache). Two rating systems were used for each station: a detailed task-specific checklist¹⁵ and a content-independent 10-point global scale. The 10-point scale is the traditional Dutch marking system, with 1 representing very insufficient performance, 6 representing (barely) sufficient performance and 10 representing excellent performance. Experienced GPs served as examiners and per station different examiners were used. The station checklist score was defined by the percentage of correctly performed items on the checklist. The station global rating score was defined by the mark of the scale. The *total test* scores were calculated by averaging across the 16 stations. The *technical stations* scores (including the nine technical stations) and the *integral stations* scores (including the seven complete patient encounter stations) were calculated by averaging across the nine technical and the seven integral stations, respectively. Based on the checklist score, the reliability (Cronbachs' alpha) was 0.63 for the total test, 0.54 for the technical stations and 0.33 for the integral stations. Based on the global rating, the reliabilities were 0.64, 0.50 and 0.33, respectively.

Variables

Eight variables were distinguished to test the three hypotheses. Two variables included the scoring on the knowledge tests, while six variables involved the scoring on the OSCE (see Table 2). The variables, expressed as mean scores, were: the knowledge test of skills score, the general knowledge test score and the OSCE total checklist and global rating score. For the OSCE four subscores were calculated: the technical stations checklist and global rating score, and the integral stations checklist and global rating score.

Procedure

The GKT was part of the national assessment programme of the eight Dutch postgraduate training institutes. The KTS and the OSCE were part of a national study into the quality of the Dutch postgraduate training in general practice.²⁷ All GP-trainees took both knowledge tests, that were administered under supervision at the institutes. The KTS took place one month before the GKT. For the OSCE test administration was centrally organised two weeks after the GKT. A national sample of 87 trainees took the OSCE. The trainees of the study presented were familiar with knowledge testing and OSCEs. The test results were used for formative assessment and research purposes.

Analysis

Per test mean and standard deviations were calculated. Pearson correlations were calculated. These observed correlations were also corrected for attenuation, assuming perfect reliability, to estimate the theoretical true relationship between variables.

Results

In Table 2 the mean test results and standard deviations are presented for all variables.

The mean test score of the KTS was the lowest (49%) and the mean test scores of the OSCE were the highest (about 68%). The differences between the various OSCE scores were minimal.

The correlations between the knowledge tests scores and the OSCE scores are shown in Table 3. The correlations between the KTS and the OSCE scores were higher than the correlations between the GKT and the OSCE scores (*hypothesis 1*). With regard to the rating system (*hypothesis 2*), the correlation between the KTS and the total OSCE was slightly higher for the checklist score than for the global rating score. In more detail, we see that for the technical stations a slightly higher correlation was found for the global rating score and that for the integral stations a considerable higher correlation was found for the checklist score.

Finally, the correlation between the KTS and the technical OSCE stations was higher than the correlation between the KTS and the integral OSCE stations (*hypothesis 3*).

Table 2

Mean scores of the KTS, GKT and OSCE, and standard deviations (SD)

Score	KTS*	GKT†	OSCE					
			Total Checklist	Total Global Rating	Technical Stations Checklist	Technical Stations Global Rating	Integral Stations Checklist	Integral Stations Global Rating
MEAN (N=47)	49.0	60.5	68.1	6.9	67.7	6.8	68.6	7.1
SD	10.7	9.0	6.8	0.5	8.8	0.7	6.6	0.6

* Knowledge Test of Skills

† General Knowledge Test

Table 3

Observed (first entry) and true (second italic entry) correlations between the KTS and the GKT*, respectively, and the OSCE

Test	OSCE											
	Total Checklist	Total Global Rating	Technical Stations Checklist	Technical Stations Global Rating	Integral Stations Checklist	Integral Stations Global Rating	Total Checklist	Total Global Rating	Technical Stations Checklist	Technical Stations Global Rating	Integral Stations Checklist	Integral Stations Global Rating
KTS	0.37	0.54	0.35	0.50	0.41	0.64	0.42	0.68	0.18	0.36	0.06	0.12
GKT	0.22	0.33	0.16	0.24	0.23	0.37	0.19	0.32	0.13	0.27	0.05	0.10

* Correlation between the KTS and GKT: 0.45 / 0.62

Discussion

In this study three hypotheses were tested concerning the predictive validity of a written knowledge test of skills in postgraduate training for general practice. Two hypotheses were confirmed. Firstly, the knowledge test of skills correlated better to the OSCE than the general knowledge test, in particular for the OSCE stations including technical, procedural skills only (*hypothesis 1*). Secondly, the knowledge test of skills correlated better to the OSCE stations including technical, procedural skills than to the OSCE including complete patient encounters (*hypothesis 3*). Finally, the knowledge test of skills correlated better to the checklist rated than to the global rated OSCE stations including complete patient encounters (*hypothesis 2*). However, we found no important difference between the checklist rated and global rated OSCE stations including technical, procedural skills. The results of this study support the predictive validity of the knowledge test of skills for an OSCE in postgraduate training for general practice.

The correlation between the knowledge test of skills and the (total) OSCE in this study was lower than the correlations reported in the literature.^{2,3,14-17} Improvement of the correlation was obtained with a better 'fit' of the knowledge test and the OSCE when including technical, procedural skills only. This probably reflects the broader skill domain that is assessed by an OSCE including complete patient encounters.

The results of this study with regard to the rating system of the OSCE warrant some further exploration. For the OSCE stations including complete patient encounters, our hypothesis has been confirmed. Apparently, the checklists of these stations correspond better to the knowledge test of skills than the global rating in which aspects such as attitude to the patient and suppleness of the consultation are judged as well. For the OSCE stations including technical, procedural skills, we found no difference between checklist and global rating, while Scherpbier and Remmen et al. found a better correlation for the checklist.^{3,17} Their OSCE also covered the more technical, procedural aspects of skills. This finding may be explained by the content of the checklists of our study, in which less detailed and more expert oriented aspects are assessed than in the checklists of undergraduate medical education. Overall, the knowledge test of skills seems more valid for the assessment of technical, procedural skills than for the assessment

of the broader domain of clinical skills used in complete patient encounters.

This study has some methodological limitations. The sample size used to estimate the true correlation was rather small. Norman et al. recommended a size of at least 100 examinees for an accurate estimation.²⁸ Another shortcoming is the fact that the tests used in this study were not real exams but formative assessments. This may influence the way examinees took the tests, although one would expect from GP-trainees at the end of postgraduate training that serious feedback is as important as passing an exam.

We conclude that, in postgraduate training for general practice, a knowledge test of skills can be used as an instrument to estimate the level of clinical skills. This is even more true when the test is used for group evaluation, such as in studies examining the efficacy of the training programme or as a screening test for deciding about courses to be offered. The more the content of the knowledge test of skills matches the clinical skills under study, the better this estimation is. Our results do not imply that written testing of knowledge of skills can replace the assessment of skills with an OSCE. Direct observation of performance of skills is more valid and stimulates trainees to practise them.

Acknowledgements

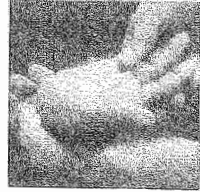
We like to thank all trainees who took part in the study. We are grateful for the support from the staff of the eight Dutch institutes for postgraduate training in general practice. We also thank the members of the national committees of the knowledge tests and the OSCE for their contribution in the development of the tests. We are grateful to IJsbrand Kramer for help with English in the writing of this paper. This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

References

1. *Lamberts H, Brouwer H, Mohrs J.* Reason for encounter, episode- and process-oriented standard output from the Transition Project. Amsterdam: Department of General Practice, University of Amsterdam; 1991.
2. *Kelly MH, Campbell LM, Murray TS.* Clinical skills assessment. *Br J General Pr* 1999;49:447-50.
3. *Scherpbier AJJA.* Kwaliteit van vaardigheidsonderwijs gemeten (Exploring the quality of skills training) [with English summary] [dissertation]. Maastricht: Universiteit Maastricht; 1997.
4. *Remmen R, Scherpbier A, van der Vleuten C, Denekens J, Derese A, Hermann I.* Effectiveness of basis clinical skills training programmes: a cross-sectional comparison of four medical schools. *Med Educ* 2001;35:121-8.
5. *Tan LHC.* Tekorten in de opleiding van huisartsen (Deficiencies in postgraduate training for general practice) [with English summary] [dissertation]. Amsterdam: Universiteit van Amsterdam; 1989.
6. *Spike N, Veitch C.* Procedural skills for general practice. *Aust Fam Physician* 1990;19:1545-53.
7. *Stillman PL, Swanson DB, Smee S et al.* Assessing clinical skills of residents with standardized patients. *Ann Intern Med* 1986;105:762-71.
8. *Van der Vleuten CPM, Swanson DB.* Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58-76.
9. *Stillman P, Swanson D, Regan MB et al.* Assessment of clinical skills of residents utilizing standardized patients. *Ann Intern Med* 1991;114:393-401.
10. *Martin JA, Regehr G, Reznick R, MacRae H, Mumaghan J, Hutchison C, Brown M.* Objective structured assessment of technical skills (OSATS) for surgical residents. *Br J Surg* 1997;84:273-8.
11. *Sloan DA, Donnelly MB, Schwartz RW, Strodel WE.* The objective structured clinical examination. *Ann Surg* 1995;222:735-42.
12. *Regehr G, Freeman R, Hodges B, Russell L.* Assessing the generalizability of OSCE measures across content domains. *Acad Med* 1999;74:1320-2.
13. *Carpenter JL.* Cost analysis of objective structured clinical examinations. *Acad Med* 1995;70:828-33.
14. *Van der Vleuten CPM, van Luyk SJ, Beckers HJM.* A written test as an alternative to performance testing. *Med Educ* 1989;23:97-107.
15. *Jansen JJM, Tan LHC, van der Vleuten CPM, van Luyk SJ, Rethans JJ, Grol RPTM.* Assessment of competence in technical clinical skills of general practitioners. *Med Educ* 1995;29:247-53.
16. *Verhoeven BH, Hamers JGHC, Scherpbier AJJA, Hoogenboom RJJ, van der Vleuten CPM.* The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. *Med Educ* 2000;34:525-9.
17. *Remmen R, Scherpbier A, Derese A et al.* Correlation of a written test of skills and a performance based test: validation study in two traditional medical schools. *Med Teacher* 2001;23:29-32.

18. *Regehr G, MacRae H, Reznick RK, Szalay D.* Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-7.
19. *Van Luyk SJ, van der Vleuten CPM.* A comparison of checklists and rating scales in performance-based testing. In: Harden IR, Hart RM, Des Marchais J, editors. Current developments in assessing clinical competence. Proceedings of the fourth Ottawa Conference. Montreal: Can.Heal Publications INC.; 1992:357-62.
20. *Cohen R, Rothman AI, Polder P, Ross J.* Validity and generalizability of global ratings in an OSCE. *Acad Med* 1991;66:545-8.
21. *Kramer AWM, Bloemen JCM, Jansen JJM.* Knowledge on skills: what is in the name? In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, editors. Advances in medical education. Dordrecht: Kluwer Academic Publishers; 1997:156-7.
22. *Grol RPTM.* National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. *Br J General Pr* 1990;40:361-4.
23. *Lamberts H, Wood M, editors.* ICPC. International Classification of Primary Care. Oxford: Oxford University Press; 1987.
24. *Jansen K.* Toetsing van technische vaardigheden van huisartsen (Assessment of technical skills of general practitioners) [with English summary] [dissertation]. Maastricht: Unigrafic; 1998.
25. *MP Springer, editor.* Basic Job Description for the general practitioner. Utrecht: Dutch National Association of General Practitioners (LHV); 1983.
26. *Van Leeuwen YD, Pollemans MC, Mol SSL, Eekhof JAH, Grol R, Drop MJ.* The Dutch knowledge test for general practice: issues of validity. *Eur J Gen Pr* 1995;1:113-7.
27. *Tan LHC, Kramer AWM.* Opleiding doorgelicht (National evaluation study of Dutch postgraduate training in general practice). Utrecht: SVUH; 1999.
28. *Norman GR, Swanson DB, Case SM.* Conceptual and methodological issues in studies comparing assessment formats. *Teach Learn Med* 1996;8:208-16.

Chapter 4



ACQUISITION OF CLINICAL SKILLS IN POSTGRADUATE TRAINING FOR GENERAL PRACTICE

This article will be published in the British Journal of General Practice as:

AWM Kramer, JJM Jansen, H Düsman, LHC Tan, CPM van der Vleuten, RPTM Grol.

Acquisition of clinical skills in postgraduate training for general practice.

Summary

Purpose

Postgraduate training in general practice aims to develop clinical competence. However, little is known about its effect on trainees' development of clinical skills. In this study (1) the acquisition of clinical skills during a three-year training programme is assessed and (2) it is evaluated whether a satisfactory level is achieved towards the end of training.

Methods

In a cross-sectional design the acquisition of clinical skills was assessed by progress-testing of knowledge of skills. The test was administered to trainees in all three years. In order to evaluate the final level, trainees at the end of the course took an OSCE and a knowledge test of skills. The results were compared to a standard of adequacy and to the results of a reference group of GP-trainers, who took the same tests.

Results

An increase in level of clinical skills and pass rate was found throughout the training, being most prominent during the first six months. At completion of their training, trainees scored higher than the GPs on the written test (48% vs. 39%) and on the OSCE (69% vs. 63%). Eighty-eight percent of the trainees would have passed the written test against 70% of the GPs and 94% would have passed the OSCE against 80% of the GPs.

Conclusion

A three-year postgraduate training in general practice contributes to a satisfactory acquisition of clinical skills. Further research into when and where acquisition takes place and the role of the GP-trainer, is needed.

Introduction

Assessment of clinical competence in postgraduate training for general practice is not only useful to evaluate trainees but may also serve to examine the quality of the training programme. Little is known about the effect of postgraduate training on the competence of trainees.¹⁻³ In a recent study we demonstrated a more or less continuous increase of general practice knowledge during Dutch postgraduate training.⁴ Although knowledge is core to any kind of clinical competence,^{1,5-7} insight in the development of other aspects of clinical competence is necessary as well for a more complete evaluation of the effects of a training programme. As clinical skills (including history taking, physical examination, diagnosing and management) constitute a crucial part of the daily work of general practitioners (GPs),⁸⁻¹⁰ it is one of the main tasks of postgraduate training in general practice to contribute to a satisfactory acquisition of these skills by trainees. Tan¹¹ and Sturmberg¹⁰ investigated whether trainees have sufficient opportunity to learn and practice clinical skills relevant for general practice during medical school and postgraduate training. The findings reported in both studies suggest that trainees, at completion of their postgraduate training, may have insufficient competence in clinical skills. However, the data were obtained by subjective ratings of trainees and trainers. More objective information about the level of clinical skills is needed to confirm this assumption.¹²

The Objective Structured Clinical Examination (OSCE) has been identified as a valid method for the assessment of clinical skills.^{12,13} However, the number of skills that can be tested in an OSCE is relatively small, due to its complex and expensive organisation. As research into clinical competence has shown a large variability between tasks and cases, assessment of a large number of skills is necessary to obtain reliable results.¹⁴ Therefore, a written test of skills can be an alternative, because it can assess many skills per test. Although the drawback of such a test is that cognitive aspects of skills are assessed rather than hands-on performance, several studies have shown that it has sufficient reliability and predictive validity for performance on an OSCE, particularly when group results are concerned as in programme evaluation.¹⁵⁻¹⁸ When combining the results of both an OSCE and a written test of skills higher levels of validity and reliability can be provided.^{19,20}

In this study we assessed the acquisition of clinical skills during a three-year postgraduate training programme in general practice and we evaluated whether a satisfactory level was achieved towards the end of training.

To assess the level of skills, a knowledge test of skills was administered to trainees in all three years and an OSCE was performed towards the end of training. The results were compared to a standard of adequacy and to the results of a reference group of GP-trainers, who took the same tests.

Methods

Context of the study

The curriculum under study was the Dutch postgraduate training in general practice from September 1995 to September 1998. Since 1988 this curriculum is organised according to a standard blueprint (see Box 1).²¹ A three years training programme was employed comprising per week four days of practical learning and one day of special training and reflection at the training institute. Eight institutes were responsible for the organisation of the training programme. The content of the programme was based on the Basic Job Description for the GP and generally aimed at the acquisition of knowledge relevant to general practice, clinical and communicative skills and attitudes. The course of the programme was structured around three blocks of one year; starting with general practice training, followed by rotation schemes in hospitals, clinics for chronically ill patients and psychiatric outpatients clinics and finished by, again, general practice training.

Instruments

National test committees using national guidelines and consensus procedures developed the knowledge test of skills and the OSCE.²² Members of the test committees were experienced GPs, with extensive teaching experience in postgraduate training, from the eight institutes and test experts.

The *knowledge test of skills (KTS)* had a progress-testing format, so that the test intended to reflect the cognitive final objectives of the curriculum.²³ It contained 180 items in the 'true/false/do not know' response format and was constructed according to a blueprint based on the International Classification for Primary Care (ICPC) and a list of clinical skills, representing a cross-section of the domain of clinical

Box 1

Dutch postgraduate training in general practice at the time of the study (1995-1998)

CONTENT			
Basic Job Description for the general practitioner			
STRUCTURE	Three blocks of equal length		Throughout the curriculum
	Block 1	General practice training	One day per week for:
	Block 2	Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics	special training and reflection, in groups of trainees, at the training institute
	Block 3	General practice training	Half-day per week for: self-directed learning
LEARNING OBJECTIVES	Block 1	Acquisition of knowledge, skills and attitudes with emphasis on common problems	
	Block 2	Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself	
	Block 3	Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations	

skills relevant to general practice. Knowledge of skills was defined as knowledge concerning the technical, procedural aspects of skills and skill-related aspects (e.g. materials and indications). The items were as much as possible embedded in vignettes representing general practice situations to increase relevancy for daily practice and assessment of crucial knowledge rather than knowledge of trivial facts. In Box 2 an example is presented.

The test score was calculated as the sum of the correct minus incorrect answers to discourage guessing (formula scoring) and was expressed as a percentage of the maximum score. The reliability (Cronbachs' alpha) of the test was 0.76, which is satisfactory, in particular for the interpretation of group results such as is used in this study.

The OSCE included 16 stations, again representing a cross-section of the clinical skills relevant for general practice, spread among 16 ICPC-categories and reflecting the final objectives of the curriculum.

Box 2

Example of two vignette-based knowledge of skills items

Mr Monisma, aged 62 years, visits his GP for an annual control of diabetes mellitus type 2 and hypertension. A physical examination of the vascular system is performed by the GP, which comprises, among other tests, the auscultation of the carotid arteries for evidence of stenosis.

- The best place for auscultation is low in the neck (just above the clavicle).

T/?/F

The GP also wants to examine the possible existence of an aneurysm of the aorta.

- It is better to examine an aneurysm of the aorta above the navel than under it.

T/?/F

Reference: Bouters EJ, Eikelboom BC, Lanson JA. Perifere vaataandoeningen. Utrecht: Wetenschappelijke Uitgeverij Bunge; 1988.

The total testing time was three hours. Nine stations involved technical, procedural tasks; in seven stations, the so-called integral stations, complete patient encounters were assessed (see Table 2). Two rating systems were used for each station: a task-specific checklist and a content-independent 10-point global scale.¹⁶ Because evidence suggests that very detailed checklists do not reflect the approach to clinical problems of the more experienced postgraduate trainee,²⁴ the checklists of our study were less detailed and more expert oriented than the checklists used in undergraduate medical education. The 10-point scale is the traditional Dutch marking system, with 1 representing very insufficient performance, 6 representing (barely) sufficient performance and 10 representing excellent performance. Eighty-four examiners were involved; all experienced GPs as well as teachers at the institutes. Most of them were familiar with the OSCE format. Each examiner was trained in the rating of two different stations. The same examiners rated the same two stations. The purpose of the training was to reach consensus in scoring between raters.

The checklist score of a station was defined by the percentage of correctly performed items on the checklist. The total test score was calculated by averaging the 16 station scores. The global rating was used for standard setting. Based on the checklist score, the reliability (Cronbachs' alpha) of the test was 0.63, which is satisfactory for the interpretations of group results such as is used in this study.

The correlation (corrected for attenuation, assuming perfect reliability) between the KTS and the OSCE was 0.54 for the total OSCE, 0.64 for the technical stations and 0.36 for the integral stations.

Procedure and subjects

The KTS was administered in spring 1998 to all Dutch GP-trainees in all three years of training (cross-sectional comparison), under supervision of the eight institutes. Six groups of trainees with training times varying from 0 to 30 months were included (see Table 1). Due to practical reasons, 775 of the total number of 1029 trainees took the test. Only for the group of trainees in their final stage of training, we were able to investigate whether participants differed from non-participants. Out of the 191 trainees of this group, 130 took the test. The participants did not differ significantly from the non-participants in age, time between graduation and start postgraduate training, entry level of general knowledge and knowledge of skills (t-test, $p > 0.05$). The percentage male of the participants was somewhat higher than the one of the none-participants (46% vs. 43%). Seven of the eight institutes were represented with a minimum of ten and a maximum of 30 trainees. In addition, 78 GP-trainers, coming from one training institute, volunteered to serve as a reference group.

One and a half months after the KTS, the OSCE was performed by a national sample of trainees in their final stage of training. Per institute a maximum of 12 were randomly selected. Seventy-three of them were included. The sample did not significantly differ from the non-participants in time between graduation and start postgraduate training and in entry level of general knowledge and knowledge of skills (t-test, $p > 0.05$). The sample was a little younger and the percentage male was somewhat higher (48% vs. 43%). Seven of the eight institutes were represented with a minimum of six and a maximum of 13 trainees. In addition, 35 GP-trainers, coming from several training institutes, volunteered to serve as reference group.

Standard setting procedures

To set a passing score for the KTS, the Cohen-Schotanus method was applied.²⁵ This absolute method uses the best performing candidates (90th percentile score) as a reference to correct for quality and test difficulty variations, under the assumption that the best performing candidates will always do well regardless of the conditions.

Subsequently, an arbitrary but conventional percentage of 55% (because this is the borderline for the Dutch marking system) of the P90 score is taken as a pass/fail point. The passing score was thus defined as the score obtained by calculating the 55% score of the P90 score (corrected for guessing) of the sample of trainees in their final stage of training.

To establish the passing score for the OSCE, the Borderline Regression method was applied; an absolute norm based on the empirical approach.²⁶ By using the global ratings of the overall performance (on the 10-point scale) the pass/fail borderline was defined at 5.5 on the scale. Per station the corresponding passing score on the checklist score scale was obtained by regressing the checklist scores on the global ratings, and then calculating the checklist score on the regression line for the global rating set at 5.5. The corresponding Borderline Regression passing score of the total test was defined by averaging the 16 station passing scores.

Analysis

For the KTS mean scores and 95% confidence intervals were calculated per group of trainees with the same training time and for the GP-trainers. Training group one comprised trainees who had just started postgraduate training, training group two comprised trainees with six months of training, training group three comprised trainees with 12 months of training, etc. A statistical significant difference was inferred from non-overlapping 95% confidence intervals. For the OSCE mean scores and standard deviations were calculated for the trainees and the GP-trainers.

The frequency of passing trainees and GP-trainers was calculated using the above-mentioned standard setting procedures.

Results

Table 1 represents the mean scores, corresponding 95% confidence intervals and pass rates of the KTS for the six training groups and the GP-trainers. The results show an increase in knowledge level of skills and pass rate over the three years of training. The increase is the most noticeable in the first six months (10.6%), whereas stagnation is observed at the end of the second year of training. The knowledge level was the highest towards the end (48.2%), although this did not differ significantly from the one attained at 18 months. Eighty-eight percent of the trainees in their final stage of training would have passed. These trainees performed better than the GP-trainers, both at the test score (48.2 % vs. 39.3 %) and the pass rate (88.1% vs. 70.5%).

Table 1

Knowledge test of skills: mean correct-incorrect score in percentage of the total, 95% confidence interval (95% CI) and pass rate for trainees and GP-trainers (cross-sectional data)

Training	Training time (months)	N	Correct-incorrect MEAN	95%CI	Pass rate (%) [*]
1	0	114	30.3	28.3 – 32.2	37.7
2	6	139	40.9	39.0 – 42.9	67.6
3	12	137	44.1	42.5 – 45.8	83.2
4	18	133	47.5	45.7 – 49.3	89.5
5	24	102	45.0	43.2 – 46.8	85.3
6 [†]	30	130	48.2	46.3 – 50.1	88.1
GP Trainers	-	78	39.3	37.3 – 41.4	70.5

^{*} passing score: 34.1 %

[†] training level at completion of training

The results of the OSCE are shown in Table 2. For the *total OSCE* the level of skills towards the end of training was 68.7%, with a pass rate of 94.5%. For the GP-trainers this level was 63.3%, with a pass rate of 80%. The mean score of the *technical stations* was somewhat lower (68.2%) than that of the *integral stations* (69.4%). Eighty-nine percent of the trainees would have passed the technical stations, while 95.1% would have passed the integral stations. For the GP-trainers the mean score of the technical stations was also lower (59.2 %) than that of the integral stations (68.6 %). Sixty percent of them would have passed the technical stations, whereas 94.3 % would have passed the integral stations.

The mean scores and pass rates of the *stations* varied considerably both between stations and groups of candidates, indicating that the competence in skills is not uniform (see Table 2).

Discussion

In this study the acquisition of clinical skills during a three-year postgraduate training programme for general practice is evaluated. Mastery of clinical skills was one of the essential learning objectives of the curriculum.

The increase in clinical skills across the years was 18% (from 30% to 48%). The greatest acquisition took place in the first sixth months. Thereafter, a rather continuous but less rapid increase is observed except for training level five, in which the increase stagnated. At this level trainees completed their second year, in which they rotated through hospitals, clinics for chronically ill patients and psychiatric outpatient clinics (see Box 1). Towards the end of training, trainees achieved a satisfactory level of both knowledge and performance of skills. They performed better than the reference group of GP-trainers and about 90% of them would have passed according to a standard of adequacy. Moreover, the final level of knowledge of skills was considerably higher than the one at medical school, as the mean final score of four medical schools at a similar knowledge test of skills varied from 18% to 39%.²⁷

Before we come to a discussion of our findings, some methodological issues have to be considered. Firstly, the acquisition of clinical skills during training was assessed by a written test of skills in a cross-sectional design.

This design requires a careful interpretation of the results. In addition,

Table 2

OSCE: mean scores in percentage of the total and standard deviations (SD), passing scores and pass rates for trainees and GP-trainers

	Correct MEAN (SD)		Passing score	Pass rate (%)	
	Trainees (N=73)	GP-trainers (N=35)		Trainees	GP-trainers
Total test	68.7 (6.1)	63.3 (7.0)	57.6	94.5	80.0
Technical stations	68.2 (8.2)	59.2 (9.3)	57.4	89.0	60.0
Insertion nasogastric tube	57.2 (17.8)	42.6 (19.5)	48.7	74.0	41.2
Resuscitation	74.5 (15.1)	57.6 (18.5)	63.3	75.3	37.1
Urinary catheterization	65.5 (18.0)	60.2 (20.6)	49.3	87.5	71.4
Painful micturition	68.6 (15.3)	61.6 (17.8)	55.5	84.7	74.3
Intravenous cannulation	77.6 (16.1)	46.8 (25.7)	59.1	83.1	32.4
Injection into the shoulder	64.3 (20.3)	68.9 (20.6)	55.9	73.6	79.4
Wound suturing	83.1 (7.5)	83.8 (7.0)	73.0	83.6	91.4
Compression bandage	55.7 (20.0)	56.6 (21.5)	52.7	52.1	60.0
Investigation vaginal discharge	67.5 (19.0)	53.6 (23.1)	59.1	61.6	31.4
Integral stations	69.4 (5.8)	68.6 (5.9)	57.8	95.9	94.3
Impaired vision	69.2 (11.7)	67.1 (14.1)	56.1	87.7	80.0
Painful knee	78.6 (11.5)	76.9 (12.3)	56.8	100.0	94.3
Fish bone throat	65.5 (20.1)	65.5 (18.8)	53.7	78.1	77.1
Backache	70.8 (11.2)	64.5 (11.1)	61.1	80.8	57.1
Pulmonary reversibility	84.7 (10.2)	81.2 (10.9)	73.0	89.0	80.0
Examination female breast	65.6 (10.3)	67.7 (14.1)	57.1	80.8	80.0
Annual control diabetes mellitus type 2	51.0 (11.3)	57.0 (10.5)	46.7	70.8	88.6

the test was not performed by all trainees in all three years. As we had no data to compare participants and non-participants, we could not control for selection bias. However, since the number of participants was substantial (about 75%), selection effects are less likely to occur. Secondly, not all trainees, at completion of their training, participated both in the written test and the OSCE. Drop out was due to practical reasons (written test) and random selection (OSCE). It so happened that only 47 trainees took both tests. For both tests participants did not differ significantly from non-participants on several aspects. Therefore and because the number of participants in both tests was relatively large, we think the results of this study are representative for the total group of trainees. Thirdly, the standard setting procedure for the knowledge test of skills is less well validated than the one for the OSCE.²⁶ The pass rates of the knowledge test are therefore less convincing than the pass rates of the OSCE. And, finally, we compared the trainees to experienced GPs who were not randomly selected out of all Dutch GPs. As they volunteered (and were thus highly motivated) and were GP-trainer as well, they probably were not representative for the Dutch GP and had more knowledge and skills than the average GP. Our findings highlight some important aspects that need further discussion. Firstly, the stagnation in growth of clinical skills at the end of the second year may possibly indicate that training in a general practice (first and third year) contributes more to the acquisition of clinical competence than training during clinical rotations (second year). This careful conclusion is supported by a study into the acquisition of knowledge during Dutch postgraduate training in general practice.⁴ In this study we also found a stagnation in growth at the end of the clinical rotations. As this study had a mixed longitudinal design, which is considered an optimal design for the measurement of change over time, its results are more robust. However, our findings do not necessarily indicate that trainees do not acquire new knowledge and skills in the second year. Perhaps our evaluation system fails to assess the acquired competence. Therefore, a closer examination of the contribution of the second year to the development of clinical competence is recommended. Secondly, the results of the OSCE show a remarkable distinction between the technical and integral stations. More trainees would have passed the integral than the technical stations, indicating that integral skills are better mastered than technical skills. Furthermore, the results of the integral stations are more or less similar for the trainees and GPs, while for the technical

stations the GPs perform less than the trainees. The implication of this finding for the teaching of technical skills needs to be further investigated. Finally, in several studies it has been found that GPs perform less than trainees at the end of training.¹ Our results are consistent with this finding. The fact that the difference is mainly found in the knowledge test of skills and the technical stations and not in the integral stations may indicate that the tests of our study are more appropriate for the assessment of trainees than experienced doctors. We conclude that a three-year postgraduate training programme for general practice, which emphasises the acquisition of clinical skills, contributes to a satisfactory acquisition of these skills. Integral skills dealing with complete patient encounters seem to be better acquired than the more technical skills. The evidence also suggests that the second year of the curriculum, in which trainees rotate through hospitals, clinics for chronically ill patients and psychiatric outpatient clinics, contributes less to the acquisition of clinical skills than the first and third year of training in a general practice. Further research into when and where acquisition of clinical skills takes place and into the role of the GP-trainer, is required. Moreover, a closer examination of the contribution of the second year to the development of clinical competence is recommended.



Acknowledgements

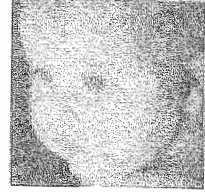
We like to thank all trainees and GPs who took part in the study. We are grateful for the support from the staff of the eight Dutch institutes for postgraduate training in general practice. We thank the members of the national committees of the KTS and the OSCE for their contribution in the development of the tests, Peter Zuithoff for his contribution in the analysis of the data and IJsbrand Kramer for improving the English. This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

References

1. *Van Leeuwen YD*. Growth in knowledge of trainees in general practice [dissertation]. Maastricht: Universitaire Pers Maastricht; 1995.
2. *Yudkowsky R*. Can resident evaluations demonstrate increase in residents' skills over time? *Acad Med* 1999;74(10Suppl):S108-10.
3. *Marel GM, Lyon PM, Field MJ, Barnsley L, Hibbert E, Parise A*. Clinical skills in early postgraduate medical trainees: patterns of acquisition of confidence and experience among junior doctor in a university teaching hospital. *Med Educ* 2000;34:1013-15.
4. *Kramer AWM, Düsman H, Tan LHC, Jansen JJM, Grol RPTM, van der Vleuten CPM*. Effect of extension of postgraduate training in general practice on the acquisition of knowledge of trainees. *Fam Pract* 2003;20:207-12.
5. *Schmidt H, Norman G, Boshuizen H*. A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990;65:611-21.
6. *Ramsey PG*. Does clinical performance correlate with physicians' scores on written examinations? In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van de Steegh AFW, editors. *Advances in Medical Education. Proceedings of the Seventh Ottawa Conference*. Dordrecht: Kluwer Academic Publishers; 1997.
7. *Ram P, van der Vleuten C, Rethans J, Schouten B, Hobma S, Grol R*. Assessment in General Practice: the predictive value of written knowledge tests and a multiple station examination for actual medical performance in daily practice. *Med Educ* 1999;33:197-203.
8. *Lamberts H, Brouwer H, Mohrs J*. Reason for encounter-, episode- and process-oriented standard output from the Transition Project. Amsterdam: Department of General Practice, University of Amsterdam; 1991.
9. *Kelly MH, Campbell LM, Murray TS*. Clinical skills assessment. *Br J General Pr* 1999;49:447-50.
10. *Sturmberg J*. Learning relevant procedural skills. *Aust Fam Physician* 1997;10:1163-5.
11. *Tan LHC*. Tekorten in de opleiding van huisartsen (Deficiencies in postgraduate training for general practice) [with English summary] [dissertation]. Amsterdam: Universiteit van Amsterdam; 1989.
12. *Friedlich M, MacRae H, Oandasan I, Tannenbaum D, Batty H, Reznick R, Regehr G*. Structured assessment of minor surgical skills (samss) for family medicine residents. *Acad Med* 2001;76:1241-6.
13. *Van der Vleuten CPM, Swanson DB*. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58-76.
14. *Van der Vleuten CPM*. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
15. *Van der Vleuten CPM, van Luyk SJ, Beckers HJM*. A written test as an alternative to performance testing. *Med Educ* 1989;23:97-107.

16. *Jansen JJM, Tan LHC, van der Vleuten CPM, Van Luyk SJ, Rethans JJ, Grol RPTM.* Assessment of competence in technical clinical skills of general practitioners. *Med Educ* 1995;29:247-53.
17. *Kramer AWM, Jansen JJM, Zuithoff P, Düsman H, Tan LHC, Grol RPTM, van der Vleuten CPM.* Predictive validity of a written knowledge test of skills in postgraduate training for general practice. *Med Educ* 2002;36:812-9.
18. *Van Dalen J, Kerkhofs E, Verwijnen GM, van Knippenberg-van den Berg BW, van de Hout HA, Scherpbier AJJA, van der Vleuten CPM.* Predicting communication skills with a paper-and-pencil test. *Med Educ* 2002;36:148-53.
19. *Swanson DB, Norman GR, Linn RL.* Performance-based assessment: lessons from the health professions. *Educational Researcher* 1995;24:5-11,35.
20. *Verhoeven BH, Hamers JGHC, Scherpbier AJJA, Hoogenboom RJJ, van der Vleuten CPM.* The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. *Med Educ* 2000;34:525-9.
21. *Dubois V, Everwijn S, Van Geldorp G et al.* The construction of a new curriculum of postgraduate training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
22. *Grol RPTM.* National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. *Br J General Pr* 1990;40:361-4.
23. *Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW.* Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teacher* 1996;18(2):103-9.
24. *Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M.* OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999;74:1129-34.
25. *Cohen-Schotanus J, van der Vleuten CPM.* Een betere cesuur bij tentamens: de beste studenten als referentiepunt (A better norm for examinations: the best students as reference point). In: Ten Cate ThJ, Dijkers JH, Houtkoop E, Pollemans MC, Pols J, Smal JA, editors. *Proceedings Gezond Onderwijs Congres* 1995. Houten/Diegem; 1996.
26. *Kramer AWM, Muijtens AMM, Jansen JJM, Düsman H, Tan LHC, van der Vleuten CPM.* Comparison of a rational and an empirical standard setting procedure for an OSCE. *Med Educ* 2003;37:132-9.
27. *Remmen R, Scherpbier AJJA, Denekens J et al.* Effectiveness of basic clinical skills training programmes: a cross-sectional comparison of four medical schools. *Med Educ* 2001;35:121-8.

Chapter 5



ACQUISITION OF COMMUNICATION SKILLS IN POSTGRADUATE TRAINING FOR GENERAL PRACTICE

This article has been submitted to Medical Education as:

AWM Kramer, H Düsman, LHC Tan, JJM Jansen, RPTM Grol, CPM Van der Vleuten.
Acquisition of communication skills in postgraduate training for general practice.

Summary

Purpose

The evidence suggests that a longitudinal training of communication skills embedded in a rich clinical context is most effective. In this study we evaluated the acquisition of communication skills under such conditions.

Methods

In a longitudinal design the communication skills of a randomly selected sample of 25 trainees of a three-year postgraduate training programme for general practice were assessed at the start and at the end of training. Per measurement and per trainee eight videotaped real life consultations were rated, using the MAAS-Global scoring list. The results were compared to each other and to those of a reference group of 94 experienced GPs.

Results

The mean score of the MAAS-Global slightly increased at the end of training (2.4) compared to the start (2.2). No significant difference was found between the final results of the trainees and the reference group. According to the criteria of the rating scale the performance of both trainees and GPs was unsatisfactory.

Conclusion

The results of this study indicate that communication skills do not improve in a three-year postgraduate training, comprising both a rich clinical context and a longitudinal training of communication skills, and that an unsatisfactory level still exists at the end of training. Moreover, GPs do not acquire communication skills during independent practice as they perform comparable to the trainees. Further research into the measurement of communication skills, the teaching procedure, the role of the GP-trainer as a model and the influence of rotations through hospitals and the like, is required.

Introduction

'Dialogue is an important means whereby the patient can acquaint the physician with those inner experiences which had led him (...) to solicit medical help'.¹ Not only this, it also influences the understanding of the health problem, the patient's satisfaction, compliance and health, and the physician's job satisfaction.²⁻⁷ So adequate communication between patient and physician is essential for good clinical practice^{8,9} and the teaching of communication skills in medical education is therefore well-established.⁷

Several studies have examined the effect of communication training in medical education and although it was found that a considerable variation exists with respect to instructional methods, content of communication skills, participants, measurement instruments and research design, some conclusions can be drawn. For instance, a significant increase in communication skills is generally found directly after the training of under- and postgraduate students as well as for physicians participating in continuous medical education.^{4-6,10-15} Long term effects are shown one to five years after completion of the training.¹⁶⁻¹⁸ When assessing communication skills throughout medical school, three studies found a decline during the clinical years (clerkships),^{3,19,20} whereas one study found the opposite.²¹ The decline was explained by the preoccupation of students with correct diagnosis and the lack of adequate communication in the hospital environment. Several studies have also assessed the impact of communication training when provided at different stages during medical education, and these studies suggested that a longitudinal training of communication skills embedded in a rich clinical context (e.g. clinical clerkships, postgraduate training) is most effective.²²⁻²⁴

This report deals with the evaluation of the acquisition of communication skills under such conditions: postgraduate training in general practice. The curriculum provides communication training throughout the entire training course, while the trainees obtain clinical experience in a general practice or hospital setting.

In order to examine the development of communication skills, real life consultations of trainees were videotaped at the start and at the end of training. The consultations were rated using the MAAS-Global scoring list.²⁵ The results were compared to each other and to those of a reference group of experienced general practitioners (GPs).²⁵



Methods

Context of the study and communication training programme

The curriculum under study was the Dutch postgraduate training in general practice from 1995 to 1998, briefly outlined in Box 1.²⁶

A three-year training programme was employed, comprising per week four days of practical learning and one day of special training and reflection at the training institute. All eight training institutes used the same programme of which its content was based on the Basic Job Description for the GP, aiming at the acquisition of knowledge relevant to general practice, clinical and communicative skills and attitudes. The course of the programme was structured around three blocks of one year; starting with general practice training, followed by rotation schemes in hospitals, clinics for chronically ill patients and psychiatric outpatient clinics and finished by, again, general practice training.

Box 1

Dutch postgraduate training in general practice at the time of the study (1995-1998)

CONTENT Basic Job Description for the general practitioner			
STRUCTURE	Three blocks of equal length	Throughout the curriculum	
	Block 1 General practice training	One day per week for:	Half-day per week for:
	Block 2 Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics	special training and reflection, in groups of trainees, at the training institute	self-directed learning
	Block 3 General practice training		
LEARNING OBJECTIVES	Block 1	Acquisition of knowledge, skills and attitudes with emphasis on common problems	
	Block 2	Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself	
	Block 3	Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations	

The training programme for communication skills was based on the principles of patient-oriented care and aimed at mastering of skills for creating a good patient-doctor relationship (communicative behaviour) and for establishing a systematic approach of the encounter (systematic behaviour). To a large extent training took place at the institute in groups of 10 trainees. About 25% of the overall training time at the institute was devoted to communication skills. In the first year a little more attention was paid to communication than in the second and third year (30% against 25%).²⁷ The educational methods comprised instruction, modelling, experiential learning (video-/audio-taping real consultations, simulations), discussion and formative assessment. Furthermore, individual attention was paid to the attitude of the trainee concerning the interaction with patients in meetings with three trainees under supervision of a trained supervisor. Apart from this, trainees worked in a general practice under supervision of a GP during the first and third year where they received modelling by and feedback of the GP-trainer.

Subjects

Seven of the eight training institutes were asked to select randomly at most eight trainees from the cohort of 191 trainees who started in September 1995 (on average 30 trainees per institute). One institute did not participate because none of their trainees started in September 1995. Due to resource restrictions, a maximum of 50 trainees for the study was feasible. The sample was tested on representativeness by personal and professional characteristics (age, gender, time between graduation and the start of postgraduate training, entree level of knowledge) (t-test and Chi-Square test). The GPs used as reference group included 25% of the GPs from the South of the Netherlands (N=94), who voluntarily participated in a study into comprehensive assessment of GPs.²⁸ The study population was representative for the Dutch GPs.

Procedure

A longitudinal design was adopted. A single cohort of trainees was assessed twice, three months after the start of their training and again after 30 months of training (six months before the end of their training). For logistic reasons a real start and end measurement was not feasible. Per measurement trainees were asked to videotape during two weeks 20 routine real life consultations.

Trainees were familiar with the videotaping of real life consultations for formative assessment purposes. For this study patients had to give informed consent before videotaping. A logbook was provided for recording consultation and patient data. Out of these 20 consultations eight consultations were selected for assessment. Research into the MAAS-Global scoring list had shown that this number of consultations yields sufficient reliability.²⁸ Selection was applied to achieve uniformity in the content of the consultations, so that comparison between the consultations of both measurements was possible. It also established a more standardised assessment situation. Criteria for selection were one reason for encounter, new illness episode and duration between three and 15 minutes. Comparability was assessed afterwards, using the selection criteria as well as gender and age of the patient, ICPC category of the patient's problem²⁹ and level of difficulty. The level of difficulty was assessed by the Amsterdam Clinical Challenge Scale (ACCS).³⁰

The consultations were rated by 27 trained GP-observers, who were familiar with the assessment instrument and postgraduate training. All rating took place after the second measurement. Each observer was sent a videotape with a, for the observer unknown, mix of individual consultations of different trainees of the first and second measurement. This procedure was used to avoid bias (i.e. the tendency to rate higher when the observer knows it is a second measurement) and to achieve uniformity in rating. The videotape also contained as much as possible consultations of different trainees to avoid a halo-effect (i.e. the tendency to rate a trainee high (or low) in all areas being evaluated if he or she scores high (or low) in one area). Finally, the videotape did not include consultations of trainees who were known to the observer, in order to achieve maximum objectivity. Two thirds of the consultations were rated by two observers to test inter-rater reliability.

The results of the second measurement were compared to those of the reference group of GPs, whose communication skills were similarly assessed in another study.²⁵

Instrument

Consultations were rated, using the MAAS-Global scoring list. This instrument reflected the learning objectives of the training programme and had an established reliability and validity.^{28,31} The list contains 11 case-independent items (see Table 2), referring to two categories of

Figure 1
Example of an item in the MAAS-Global scoring list

FLEXIBILITY	0	1	2	3	4	5	6*
Empathising, attentive and open in intonation, gesture and eye contact							
Adequate time/space for the patient							
No disrupting hitches or interruptions							

* 0=absent, 1=bad, 2=insufficient, 3=doubtful, 4=sufficient, 5=good, 6=excellent

communication skills: skills needed in certain phases of the consultation (e.g. opening) and general skills needed in the entire consultation (e.g. dealing with emotions). Items are global but anchored with detailed criteria. They are scored on a 7-point Likert scale, ranging from 'not present' to 'excellent' (see Figure 1 for an example). All items must be completed in each consultation. The reliability using generalisability theory was 0.71, which is satisfactory, in particular for the interpretation of group results such as is used in this study.

Analysis

Per trainee two scores were calculated: an item score for each of the 11 MAAS-items and a total test score. The item score was calculated by averaging across the eight consultations. First calculating the mean score per consultation and then averaging across the eight consultations established the total test score. To calculate per item the mean item score and standard deviation, item scores were averaged across the number of trainees and GPs, respectively. For the mean total test score and standard deviation, the total test scores were averaged across the same groups. Mean items and total test scores of the first and second measurement were compared and differences were tested on statistical significance with a paired t-test. Mean items and total test scores of the second measurement and the GPs were also compared and differences were tested on statistical significance with a t-test.



Results

Subjects

Thirty-nine trainees participated in the first measurement, at the start of the training course. Of these, only 25 participated at the second measurement, at the end of the three years training. The 25 trainees did not significantly differ from the 166 non-participants with respect to the controlled aspects as depicted in Table 1. However, one of the seven participating institutes was not represented.

Comparability of the consultations

The consultations of the first and second measurement were comparable with respect to all controlled aspects as described in the methods section.

MAAS-Global scoring list

Table 2 presents the mean scores on the 11 MAAS-items and the total test, for both trainees and GPs. When comparing the performance of the trainees on the two measurements, the results show hardly any difference.

Table 1

Comparison between personal and professional characteristics of the participants and the non-participants

Characteristics	Participants (N=25)	Non-participants (N=166)	Significant difference
	MEAN	MEAN	
Age (years)	30	30.6	n.s.
Gender (% female)	48	56	n.s.
Time between graduation and start postgraduate training (months)	30	29.7	n.s.
Start level general knowledge (%)	38.2	39.1	n.s.
Start level knowledge of skills (%)	27.9	26.1	n.s.

Table 2

Mean MAAS-Global scores per item and total scores of the trainees and the GPs (scoring scale 0 - 6⁺)

Items	Trainees (N=25)		GPs (N=94)
	Start of training MEAN (SD)	End of training MEAN (SD)	MEAN (SD)
Skills per phase			
Opening	2.4 (0.6)	2.7 (0.6)	2.8 (0.6)
Clarification of the reason for encounter	1.4 (0.6)	1.5 (0.7)	1.4 (0.7)
Communication about the diagnosis/hypothesis	3.2 (0.6)	3.3 (0.6)	3.1 (0.6)
Communication about the management plan	2.9 (0.5)	2.9 (0.6)**	2.6 (0.5)**
Evaluation of the consultation	0.7 (0.4)	0.8 (0.5)**	1.4 (0.5)**
General skills			
Providing information	2.8 (0.5)	2.8 (0.4)	2.9 (0.5)
Dealing with emotions	0.9 (0.5)	1.0 (0.5)	1.0 (0.5)
Summarising	1.5 (0.5)†	1.7 (0.6)†	1.7 (0.8)
Ordering	3.2 (0.5)	3.4 (0.6)	3.5 (0.6)
Flexibility in the communication	3.6 (0.6)†	3.9 (0.5)†	4.0 (0.6)
Exploration of expectations and feelings	1.4 (0.7)	1.7 (0.5)	Not assessed
Total	2.2 (0.4)†	2.4 (0.4)†	2.4 (0.4)

* 0 = absent, 1 = bad, 2 = insufficient, 3 = doubtful, 4 = sufficient, 5 = good, 6 = excellent

† p < 0.05 (paired t-test)

** p < 0.05 (t-test)

Only the scores at the items 'summarising' and 'flexibility in the communication with the patient' were significantly higher at the second measurement, but differences were small. With regard to the total test, the mean score slightly improved at the end of training (2.4) compared to the start (2.2). The highest scores were found on communication skills dealing with diagnosis, ordering and flexibility (scores > 3). The lowest scores were found on communication skills dealing with clarification of the reason for encounter, evaluation of the consultation, emotions, summarising and exploration of expectations and feelings (scores < 2).

According to the criteria of the scoring scale, the performance of the trainees towards the end of training was unsatisfactory or doubtful for all items except 'flexibility' that, with a score of 3.9, approximated a satisfactory score. The overall performance was unsatisfactory. In the comparison between the scores of the trainees at the end of training and the experienced GPs, again little difference was found. Only for 'communication about the management plan' and 'evaluation of the consultation', some significant difference was shown.



Discussion

In this study we tested the hypothesis that a postgraduate training in general practice, comprising both special, longitudinal communication training and a rich clinical context, would favour the acquisition of communication skills of trainees. This hypothesis could not be confirmed. Although some increase in overall communication skills was found, this increase was minimal in view of a 27 months training period. Moreover, the overall level was unsatisfactory. With regard to the scoring at the individual items of the MAAS-Global list, two patterns can be distinguished. Firstly, the level of some communication skills was already relatively high at the first measurement (opening, communication about diagnosis and management plan, providing information, ordering and flexibility). These are the more traditional aspects of patient - doctor interaction (corresponding to history taking, physical examination, diagnosis and treatment, and aspects of routine and structure). Apparently, trainees have learnt these skills to some extent at undergraduate education. The level of these skills is, however, still unsatisfactory and one would have expected it to increase during postgraduate training. Secondly, and remarkably, for the more basic communication skills (clarification of the reason for

encounter, evaluation of the consultation, dealing with emotions, summarising and exploration of expectations and feelings) the results show a low score both at the start and towards the end of training with the exception of 'summarising', where a small but significant improvement was found. This is all the more surprising, because important aspects such as 'clarification of the reason for encounter' and 'exploration of expectations and feelings' received great emphasis during the training but apparently without noticeable effect. Both the lack of acquisition of communication skills and its unsatisfactory final level indicate a remarkable shortcoming of postgraduate training in general practice. The important question it raises, is whether or not such skills are subsequently acquired during independent clinical practice. The results of the reference group of experienced GPs suggest that GPs do also not acquire communication skills, as the reference group performed more or less comparable to the trainees. For a discussion of these findings successively three aspects will be considered: the design of the study, the instrument and the training of communication skills.

Design of the study

Some limitations in the design of the study exist. The population under study was rather small although representative. Participation was also voluntary, which favours selection of good performers rather than bad performers. Another shortcoming is that the first measurement was not a real start measurement. Perhaps the increase would have been greater with a real start measurement. Research into the effect of postgraduate training on the acquisition of knowledge and skills shows the greatest growth in the first six months.^{32,33} Still, 27 months remained for improvement. Finally, it is recommended in the literature to measure communications skills on different levels including knowledge, attitude, competence, actual performance and patient outcome.^{3,34} It is possible that with different measures more training effect would have been found. Moreover, in our study we have assessed communication skills on the 'does' level, according to the theoretical concept of clinical assessment as devised by Miller.³⁵ This implies that the extent to which trainees and GPs put into practice their communication skills is tested rather than their capacity to communicate. Perhaps a greater increase in communication skills would have been shown when communication skills were assessed in, for instance, an Objective Structured Clinical Examination (OSCE).

The instrument

The instrument to measure performance may have shortcomings. The MAAS-global list is a case-independent instrument. Research into medical competence has shown that competence is highly content related, but that this is less dramatic for communication.²⁶ Thomson et al. found that performance on similar cases, such as in our study, is modestly comparable.³⁶ Still, the question remains whether all communication skills of the MAAS-global list are required in each general practice encounter. Fraser et al. reported a study into the validity and reliability of the Leicester assessment package, an instrument to measure competence in general practice consultation. They allow examiners flexibility to adjust their construct of competence to match the particular challenges posed in each consultation.³⁷ Furthermore, although the MAAS-global list generally reflects the learning objectives of the communication training, it possibly does not reflect the objectives of all individual trainers and trainees. There is a lot of discussion about adequate communication in general practice. Opinions about good communication change during the time, due to changes in society and GP's tasks.³⁸ In contrast with medical knowledge and skills, which are more evidence based and less open for discussion, communication is more affected by personal convictions and attitudes. This may reduce the chance of finding training effects.⁸ But on the other hand, the reliability was high, indicating that the repeated measurements across consultations rank ordered the trainees consistently. As different raters rated different consultations, this can be no spurious effect of a possible halo effect across raters. Different consultations apparently 'agreed' on the rank ordering of trainees in terms of their communication ability.

Communication skills training

The communication training may have had shortcomings. With regard to the training at the training institutes, it seems that appropriate educational methods were used.²³ Moreover, the trainees appreciated the communication training the most of all special education they received.²⁷ They also felt more confident in performing communication skills at the end of training than at the start of it.³⁸ Perhaps communication needs more attention in the second year, in which trainees rotate through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. Williams et al. found that preregistration house officers improved their communication skills in general practice,

but that it was difficult for them to bring their improved skills back into the hospital setting.³⁹ Also patient feedback did not occur as education tool, while Greco et al. found that this was an effective intervention with regard to improving communication skills.²⁴

Apart from special communication training, a significant role was reserved for the GP-trainer as model. Research has shown that Dutch experienced GPs and GP-trainers do not always put in practice the communication skills that have to be learnt during postgraduate training.^{25,41,42} Therefore, their model function may have been less adequate.

Conclusion

The results of this study suggest that communication skills do not improve in a three-year postgraduate training programme, comprising both a rich clinical context and a special, longitudinal communication training, as was recommended in the literature. Further research into the measurement of communication skills, the teaching procedures, the role of the GP-trainer as a model and the influence of rotations through hospitals and the like, is required to lift a greater corner of the veil over this important but complex domain of general practice.



Acknowledgements

We like to thank all the trainees who took part in the study. We are grateful for the support from the GP-trainers and the staff of the participating institutes for postgraduate training in general practice. We also thank Paul Ram for providing the scores of the experienced GPs and for training the observers, and Peter Zuithoff for his contribution in the data analysis. We are grateful to IJsbrand Kramer for help with English in the writing of this paper. This study was initiated by the Registration Committee for Postgraduate Training in General Practice (HVRG), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

References

1. *Engel GL*. How much longer must medicine's science be bound by a seventeenth century world view? In: White KL, editor. The task of medicine. Dialogue at Wickenburg. Menlo Park, Ca: The Henry J. Kaiser Family Foundation; 1988.
2. *Stewart MA*. Effective physician-patient communication and health outcomes: a review. *Can Med Assoc J* 1995;152:1423-33.
3. *Pfeiffer C, Madray H, Ardolino A, Willms J*. The rise and fall of students' skill in obtaining medical history. *Med Educ* 1998;32:283-8.
4. *Winefield HR, Chur-Hansen A*. Evaluating the outcome of communication skill teaching for entry-level medical students: does knowledge of empathy increase? *Med Educ* 2000;34:90-4.
5. *Langewitz WA, Eich P, Kiss A, Wössmer B*. Improving communication skills- a randomized controlled behaviourally oriented intervention study for residents in internal medicine. *Psychosom Med* 1998;60:268-76.
6. *Hulsman RL, Ros WJG, Winnubst JAM, Bensing JM*. Teaching clinically experienced physicians communication skills. A review of evaluation studies. *Med Educ* 1999;33:655-68.
7. *Spencer J, Silverman J*. Education for communication: much already known, so much more to understand. *Med Educ* 2001;35:188-90.
8. *World Health Organization*. Doctor-patient interaction and communication. Genova: WHO; 1993.
9. *Walton HJ, editor*. World Federation for Medical Education. Proceedings for the World Summit on Medical Education. *Med Educ* 1993;28(1Suppl.):40-9.
10. *Carroll JG, Monroe J*. Teaching clinical interviewing in the health professions. A review of empirical research. *Eval Health Prof* 1980;3:21-45.
11. *Grol R, Mokkink H, Helsper-Lucas A, Thielens V, Bulte J*. Effects of the vocational training of general practice on consultation skills and medical performance. *Med Educ* 1989;23:512-21.
12. *Roter DL, Hall JA, Kern DE, Barker LR, Cole KA, Roca RP*. Improving physicians' interviewing skills and reducing patients' emotional distress. A randomized clinical trial. *Arch Intern Med* 1995;155:1877-84.
13. *Bensing JM, Sluijs EM*. Evaluation of an interview course for general practitioners. *Soc Sci Med* 1995;20(7):737-44.
14. *Bögels SM*. Immediate and long-term effects of psychiatric interview training on students' skills. *Med Teacher* 1996;18:279-87.
15. *Farnhill D, Todisco J, Hayes SC, Bartlett D*. Videotaped interviewing of non-English speakers: training for medical students with volunteer clients. *Med Educ* 1997;31:87-93.
16. *Maguire P, Fairbairn S, Fletcher C*. Consultation skills of young doctors: benefits of feedback training in interviewing as students persist. *BMJ* 1986;292:1573-6.
17. *Rudner HL, Bestvater D, Bader E*. Evaluating family counselling skills training for family practice. *Med Educ* 1990;24:461-6.
18. *Oh J, Segal R, Gordon J, Boal J, Jotkowitz A*. Retention and use of patient-centered interviewing skills after intensive training. *Acad Med* 2001;76:647-50.



19. *Kraan HF, Crijnen AAM, de Vries MW, Zuidweg J, Imbos T, van der Vleuten CPM.* To what extent are medical interviewing skills teachable? *Med Teacher* 1990;12:315-28.
20. *Craig JL.* Retention of interviewing skills learned by first-year medical students: a longitudinal study. *Med Educ* 1992;26:276-81.
21. *Klamen DL, Williams RG.* The effect of medical education on students' patient-satisfaction ratings. *Acad Med* 1997;72:57-61.
22. *Van Dalen J, Kerkhofs E, van Knippenberg-van den Berg BW, van den Hout HA, Scherpbier AJJA, van der Vleuten CPM.* Longitudinal and concentrated communication skills programmes: two Dutch medical schools compared. *Adv Health Sci Educ* 2002;7:29-40.
23. *Aspegren K.* BEME Guide No.2: Teaching and learning communication skills in medicine-a review with quality grading of articles. *Med Teacher* 1999;21:563-70.
24. *Greco M, Brownlea A, McGovern J.* Impact of patient feedback on the interpersonal skills of general practice registrars: results of a longitudinal study. *Med Educ* 2001;35:748-56.
25. *Ram R, Grol R, Rethans J, Schouten B, van der Vleuten C, Kester A.* Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Med Educ* 1999;33:447-54.
26. *Dubois V, Everwijn S, van Geldorp G et al.* The construction of a new curriculum of postgraduate training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
27. *Tan LHC, Kramer AWM.* Opleiding doorgelicht (National evaluation study of Dutch postgraduate training for general practice). Utrecht: SVUH; 1999.
28. *Van Thiel J, van der Vleuten C, Kraan H.* Assessment of medical interviewing skills: generalizability of scores using successive MAAS-versions. In: Harden RM, Hart IR, Mulholland H, editors. *Approaches to the assessment of clinical competence.* Dundee: Centre for medical education; 1992: 536-40.
29. *Lamberts H, Wood M, editors.* *ICPC.* International Classification of Primary Care. Oxford: University Press; 1987.
30. *Gercama AJ, De Haan M, van der Vleuten CPM.* Reliability of the Amsterdam Clinical Challenge scale (ACCS): a new instrument to assess the level of difficulty of patient cases in medical education. *Med Educ* 2000;34:519-24.
31. *Van Dalen J, Prince CJAH, Scherpbier AJJA, van der Vleuten CPM.* Evaluating communication skills. *Adv Health Sci Educ* 1998;3:187-95.
32. *Kramer AWM, Düsman H, Tan LHC, Jansen JJM, van der Vleuten CPM, Grol RPTM.* Effect of extension postgraduate training in general practice on the acquisition of knowledge of trainees. *Fam Pract* 2003;20:207-12.
33. *Kramer AWM, Jansen JJM, Düsman H, Tan LHC, van der Vleuten CPM, Grol RPTM.* Acquisition of clinical skills in postgraduate training for general practice. Accepted for publication in *Br J General Pract.*
34. *Greco M, Spike N, Powell R, Brownlea A.* Assessing communication skills of GP registrars: a comparison of patient and GP examiner ratings. *Med Educ* 2002;36:366-76.



35. *Miller GE.* The assessment of clinical skills/competence/performance. *Acad Med* 1990;65 (9Suppl):S63-7.
36. *Thomson AN.* Can communication skills be assessed independently of their context? *Med Educ* 1992;26:364-7.
37. *Fraser RC, McKinley RK, Mulholland H.* Consultation competence in general practice: testing the reliability of the Leicester assessment package. *Br J General Pr* 1994;44:293-6
38. *Mol A.* Praten met patienten. Een geschiedenis [Talk with patients. A history]. *Huisarts Wet* 1991;34: 306-10.
39. *Kramer AWM, Zuithoff P, Jansen JJM, Tan LHC, Grol RPTM, van der Vleuten CPM.* Growth of self-perceived clinical competence in postgraduate training for general practice. Submitted.
40. *Williams C, Cantillon P, Cochrane M.* The doctor-patient relationship: from undergraduate assumptions to pre-registration reality. *Med Educ* 2001;35:743-7.
41. *Vernooij-Dassen MJFJ, Ram PM, Brennikmeijer WJM et al.* Quality assessment in general practice trainers. *Med Educ* 2000;34:1001-6.
42. *Bügel PC, Meyboom- De Jong B, Roordink MH, van de Ven NJAM, Groenier KH.* Geruuststellen in de huisartspraktijk [Reassure in general practice]. *Huisarts Wet* 2000;43:559-62.



Chapter 6



GROWTH OF SELF-PERCEIVED CLINICAL COMPETENCE IN POSTGRADUATE TRAINING FOR GENERAL PRACTICE

This article will be submitted as:

AWM Krarner, P Zuithoff, JJM Jansen, LHC Tan, RPTM Grol, CPM van der Vleuten.

Growth of self-perceived clinical competence in postgraduate training for general practice.

Summary

Purpose

Self-perceived clinical competence is an important aspect of medical education, in particular for postgraduate training. The purpose of this study was (1) to examine the increase in self-perceived clinical competence during a three-year postgraduate training in general practice and (2) to explore the relation between the growth of self-perceived competence and several background variables. The training programme consisted of learning in a rich clinical environment as well as additional special education.

Methods

In a longitudinal design all Dutch trainees of one cohort (N=191) were asked to complete a questionnaire, which assessed their self-perceived knowledge, clinical skills and consultations skills, at the start and at the end of the course. Information about potentially influencing background variables was also collected. Amongst these were variables such as: age, gender, prior medical experience, the effort someone has spent upon her/his education, insight in weak and strong areas of clinical competence and knowledge and skills levels.

Results

127 trainees completed both questionnaires. Growth of self-perceived clinical competence was found throughout the three years of training. A satisfactory level was achieved at the end of the course. Self-perceived consultation skills increased more than self-perceived knowledge and clinical skills. The afore mentioned background variables did not relate in any way with the growth of self-perceived clinical competence.

Conclusion

This study shows that a three-year postgraduate training in general practice enhances self-perceived clinical competence. However, we still do not know how to explain this improvement. Further study into the theoretical concept of self-assessment in medical education and into the factors contributing to the feeling of being competent, is required.

Introduction

The value of self-assessment of clinical competence as part of the individual learning process is obvious.^{1,2} Many authors consider it a skill of learners or professionals that enables a person to assess one's own knowledge and skills; and like other skills this skill has also to be learned.³ They emphasise the importance of this skill for self-directed learning as a basis for lifelong maintenance of competence and for professional autonomy.³⁻⁶ However, the value of self-assessment for the purpose of evaluation of clinical competence is less clear.^{1,2} Validity and accuracy of self-assessment used for this purpose have been extensively investigated in medical education. Although substantial variation in study design existed, all studies attempted to validate self-assessment by comparing the results of self-assessment to those of expert ratings or objective tests. Research evidence has provided little support for the validity of self-assessment in relation to external evaluations.^{4,7-19} Even repeated personal feedback based on objective knowledge and skills scores did not improve the accuracy.¹⁹ From these findings it may be concluded that self-assessment of clinical competence apparently reveals to what extent someone considers her/himself competent (self-perceived competence)²³⁻²⁵ rather than actual clinical competence.^{2,16,19,21,22} It is obvious that from a medical education programme it may be expected that it not only will support the transfer of knowledge and skills but also the extent to which trainees consider themselves competent. This is especially applied for a postgraduate training that prepares trainees to bare formal responsibility for patient care. Moreover, a postgraduate training often comprises both a rich clinical context (general practice or hospital setting) and additional special educational programmes. This combination may give trainees a good opportunity to increase their sense of competence. In the present study we examined the effect of a postgraduate training programme for general practice on the development of self-perceived clinical competence. We expected to find an increase of it. Because little is known about predictors of self-perceived clinical competence, we also explored the relation between the growth of self-perceived competence and several background variables that may potentially be of influence to it. The following variables were recognised: age, gender, prior medical experience, the effort someone has spent upon her/his education and insight in weak and strong areas of clinical competence.

Furthermore, although a direct relation between self-perceived and actual clinical competence seems not to exist, knowledge and skills may influence the growth of self-perceived competence. Therefore, we have also explored the relation between knowledge test scores and growth of self-perceived competence.

Methods

Procedure and subjects

A longitudinal design was adopted. All Dutch trainees (N=191) who started their postgraduate training in September 1995 were asked to take part in the study. Participation included the completion of a questionnaire, which assessed their self-perceived clinical competence, at the start and at the end of their postgraduate training. Information about potentially influencing background variables was collected among them as well. The results of both measurements were compared to each other, and the relative contribution of the potentially influencing background variables to the outcome was analysed.

Context of the study

The curriculum under study was the Dutch postgraduate training in general practice from 1995 to 1998, briefly outlined in Box 1.²⁶ A three years training programme was employed comprising per week four days for practical learning and one day for special training and reflection at the training institute. Eight institutes were involved in the organisation of the training. The content of the programme was based on the Basic Job Description for the general practitioner (GP) and generally aimed at the acquisition of knowledge relevant to general practice, skills and attitudes. The course of the programme was structured around three blocks of one year; starting with general practice training, followed by rotation schemes in hospitals, clinics for chronically ill patients and psychiatric outpatients clinics and finished by, again, general practice training.

Instrument

A self-assessment questionnaire was constructed according to the learning objectives of the training programme: the acquisition of knowledge and skills. For skills a distinction was drawn between clinical skills and consultation skills.

Box 1

Dutch postgraduate training in general practice at the time of the study (1995-1998)

CONTENT	Basic Job Description for the general practitioner		
STRUCTURE	Three blocks of equal length	Throughout the curriculum	
	Block 1 General practice training	One day per week for:	Half-day per week for:
	Block 2 Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics	special training and reflection, in groups of trainees, at the training institute	self-directed learning
	Block 3 General practice training		
LEARNING OBJECTIVES	Block 1 Acquisition of knowledge, skills and attitudes with emphasis on common problems Block 2 Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself Block 3 Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations		

The domain of *knowledge* was represented by the International Classification of Primary Care (ICPC).²⁷ For the 17 categories of the ICPC trainees were asked to estimate their knowledge level on a 3-point scale (none, some, considerable knowledge).

The domain of *clinical skills* was represented by both the ICPC and a list of specific patient groups (children, the elderly, working population, migrants, violence victims, addicts, patients with acute or chronic diseases, dying persons). The trainees were asked to estimate their level of clinical skills on a 4-point scale (hardly any, moderate, reasonable, good mastery).

The domain of *consultation skills* was represented by a list of five items, summarising the main aspects of the consultation process (systematic approach of the encounter, diagnostic and therapeutic management, maintenance of a good patient-doctor relationship, support of the continuity of care and preventive activities).

The trainees were asked to estimate their level of consultation skills on a 4-point scale (hardly any, moderate, reasonable, good mastery). The reliability (Cronbachs' alpha) of the start and end measurement was 0.8 / 0.9 for knowledge, 0.8 / 0.9 for clinical skills and 0.7 / 0.7 for consultation skills, respectively. These are satisfactory reliabilities, particularly for the interpretation of group results as is required in this study.

Factors influencing self-perceived clinical competence

Based on the literature and on common sense, the following factors were recognised as potentially influencing growth of self-perceived clinical competence.

Age and prior medical experience were chosen, because they reflect (life) experience that may result in increased self-confidence. Bleys et al. found that prior medical experience has a positive effect on self-perceived growth in some aspects of competence during a short training programme.²⁸ Prior medical experience was defined as the time between graduation and the start of postgraduate training.

Gender has been identified as a factor that may influence self-assessment. Women are perhaps more self-critical than men, resulting in less self-perceived competence.²⁹

Although the main content, structure and learning objectives of the training programme were similar for the eight *training institutes*, variations existed between the ways this programme was performed at the individual institutes. This may have lead to differences in self-perceived clinical competence of trainees.

Personal effort has been recognised as a factor that may influence self-assessment: the greater the effort someone put in her/his medical education, the higher the self-assessment of clinical competence.¹¹

Therefore, we asked the trainees, at completion of their training, to rate the effort they had spent upon their postgraduate training.

Furthermore, we asked the trainees to what extent teachers had give them *insight* in their weak and strong areas throughout the course. This insight is expected to result in more growth of self-perceived competence.

Finally, the start and final *knowledge test scores* of the participating trainees were used as background variables, to examine the influence of knowledge and clinical skills on the growth of self-perceived clinical competence. In Dutch postgraduate training knowledge tests are regularly administered. These tests have a progress-testing format, so

that each test has the intention to reflect the cognitive final objectives of the curriculum. To enhance general practice relevance, items are embedded in vignettes representing real consultations. Two different tests exist: a General Knowledge Test (GKT) and a Knowledge Test of Skills (KTS).^{30,31} For the KTS research has shown that it can predict performance of clinical skills in an OSCE.³⁰

Analysis

Per measurement moment, mean scores and standard deviations were calculated for the domains of self-perceived knowledge, clinical skills and consultation skills. Differences were tested on statistical significance (paired t-test).

In order to explore the relation between the growth of self-perceived clinical competence and the potentially influencing background variables (age, gender, prior medical experience, effort, insight, knowledge test scores), partial correlations were calculated between these variables and the final self-assessment score (post-test), controlling for the self-assessment score at the start of training (pre-test), thereby introducing 'growth' as the actual dependent variable.³² For the influence of the training institute on growth, a General Linear Model (GLM) analysis was applied.³³ To investigate whether a combination of potentially influencing background variables could better explain the growth of self-perceived competence, a multiple regression analysis was performed. Moreover, in a regression analysis the growth of knowledge and skills during the course was introduced by controlling the knowledge test scores on the end measurement for those on the start measurement.

Results

Subjects

Out of 191 trainees, 127 trainees completed both self-assessment questionnaires. The responders did not significantly differ from the non-responders with respect to age, time between graduation and the start of postgraduate training, entry level of general knowledge and knowledge of skills (t-test, $p > 0.05$). The percentage male trainees of the responders was a little higher than the one of the non-responders (47% versus 41%). Seven of the eight Dutch training institutes were represented with a minimum of four and a maximum of 34 trainees.

Table 1

Mean scores and standard deviations (SD) of the domains knowledge, clinical skills and consultation skills of the self-assessment questionnaire

Domain	N	Start of training MEAN (SD)	End of training MEAN (SD)	Significance*
Knowledge (3-point scale) [†]	123	2.0 (0.2)	2.5 (0.3)	P<0.001
Clinical skills (4-point scale)**	125	2.5 (0.4)	3.2 (0.3)	P<0.001
Consultation skills (4-point scale)**	127	2.1 (0.4)	3.4 (0.4)	P<0.001

* t-test

[†] 1 = none, 2 = some, 3 = considerable

** 1 = hardly any, 2 = moderate, 3 = reasonable, 4 = good

Self-assessment questionnaire

In Table 1 the results per domain of self-perceived clinical competence are presented.

Self-perceived knowledge increased from 'some knowledge' to a level between 'some-' and 'considerable knowledge'. For self-perceived clinical skills trainees started with a level between 'moderate' and 'reasonable' mastery and it increased to 'more-than-reasonable' mastery. Self-perceived consultation skills showed the greatest increase, from 'moderate' mastery to a level between 'reasonable' and 'good' mastery. All differences were statistically significant.

Background variables

Background variables were normally distributed. Table 2 shows the partial correlations, representing the relation of the variables age, prior medical experience, gender, effort, insight and knowledge test scores with the growth of self-perceived knowledge, clinical skills and consultation skills.

For none of these background variables a relation with growth in self-perceived competence was demonstrated. Training institute as background variable also showed no influence on the growth of self-perceived competence.

Table 2

Partial correlations between self-perceived knowledge, clinical skills and consultation skills, respectively, and the potentially influencing factors

Potentially influencing background variables	Self-perceived knowledge		Self-perceived clinical skills		Self-perceived consultation skills	
	Corr	N	Corr	N	Corr	N
Age	-0.07	118	0.02	120	-0.07	122
Prior medical experience	-0.07	120	0.01	122	-0.06	124
Gender	-0.09	120	0.12	122	-0.03	124
Effort	-0.10	119	-0.16	121	-0.10	123
Insight	0.04	120	0.04	122	0.05	124
GKT* start measurement	-0.03	120	0.0	122	-0.04	124
GKT* end measurement	0.19	107	0.10	109	0.1	111
KTS† start measurement	0.01	120	0.07	122	-0.05	124
KTS† end measurement	0.07	85	-0.07	87	-0.1	89

* General Knowledge Test

† Knowledge Test of Skills

The regression analyses yielded no evidence for a significant contribution of the chosen variables to the growth in self-perceived knowledge, clinical and consultation skills, respectively. For the three domains less than 25% of the variance was explained by all variables in combination.

Discussion

The results of this study show that a three-year postgraduate training in general practice, comprising both learning in a clinical setting and additional special education, improved the extent to which trainees

considered themselves competent in performing knowledge relevant to general practice, clinical skills and consultation skills. Moreover, they felt themselves sufficiently competent on all aspects at the end of the course. With regard to the three examined domains of clinical competence, a somewhat different 'growth' pattern was found. At the start of the course trainees considered themselves more or less as competent in knowledge as in clinical skills, whereas they felt less competent in consultation skills. At the end this was the opposite. Apparently, the training programme enhanced the self-perceived competence in consultation skills more than that in knowledge and clinical skills.

Our second research question was whether or not background variables could be found that explained the growth of self-perceived competence. With regard to the selected variables (age, prior medical experience, gender, training institute, the effort someone has spent upon her/his education, insight in weak and strong areas of clinical competence and knowledge test scores), the results of this study did not provide conclusive evidence.

The overall result of the first part of our study was consistent with what we assumed: postgraduate training supported self-perceived clinical competence. Because we did not have special expectations with regard to the three distinguished domains of clinical competence, the difference in growth between them was unexpected.

Apart from this study, we examined, in three other studies with the same cohort of trainees, the acquisition of actual knowledge, clinical skills and communication skills.^{30,34,35} These aspects of clinical competence were assessed by general knowledge tests, knowledge tests of skills and the MAAS-Global scoring list for communication skills, respectively. In these studies an increase of knowledge and clinical skills was shown and no improvement of communication skills. These results support the finding of the second part of our study; namely that an increase in objectively measured clinical competence does not explain the gain in self-perceived clinical competence. Apparently, other factors than an increase of 'real' competence enhance the feeling of being competent. Unfortunately, our study does not seem to reveal the identity of such factors, as the other examined background variables also did not explain the growth of self-perceived competence. For instance, it is surprising that insight in weak and strong areas of clinical competence or gender were not related to an increase of self-perceived competence. Our findings extend the

confusion about what is measured by self-assessment of competence. In the literature several aspects have been mentioned that may refer to the concept of it (self-confidence²³, potential or ideal performance^{11,13}, effort^{11,13}, defensiveness⁷, maintenance of self-esteem¹³, non-cognitive competence⁹ and self-attributions⁴). Although these interpretations seem plausible, we have not found a publication in which the theoretical concept of self-assessment in medical education has been explained profoundly. As self-assessment of competence is widely used as an assessment procedure in medical education and the importance of it for self-directed learning and professional autonomy is well established,^{3,6} the need for such a theoretical explanation is urgent.

In conclusion, we have shown in this study that a three-year postgraduate training in general practice enhances the self-perceived competence in knowledge, clinical skills and consultation skills. However we still do not know how this improvement is explained. More study into the theoretical concept of self-assessment and into the factors contributing to the feeling of being competent, is required.

Acknowledgements



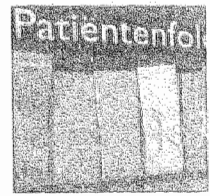
We like to thank all trainees who took part in the study. We are grateful for the support from the staff of the eight Dutch institutes for postgraduate training in general practice. We thank the members of the national committees of the GKT and the KTS for their contribution in the development of the tests. Herman Düsman contributed in the design of the study, the administration and the analysis. IJsbrand Kramer improved the English. This study was initiated by the Registration Committee for Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH) and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

References

1. *Boud D.* The role of self-assessment in student grading. *Assessment and evaluation in higher education* 1989;14:20-30.
2. *Purdy M.* The problem of self-assessment in nurse education. *Nurse Education Today* 1997;17:135-9.
3. *Miller GE.* Continuous assessment. *Med Educ* 1976;10:81-6.
4. *Gordon MJ.* A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991;66:762-9.
5. *Irvine D.* The performance of doctors. I: Professionalism and self regulation in a changing world. *BMJ* 1997;314: 1540-2.
6. *Oermann MH.* Professional reflection: have you looked in the mirror lately? *Orthopaedic Nursing*;1998:22-6.
7. *Harris MH, Schaubroeck J.* A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology* 1988;41:43-62.
8. *Stuart MR, Goldstein HS, Snope FC.* Self-evaluation by residents in family medicine. *J Fam Pract* 1980;10:639-42.
9. *Arnold L, Willoughby TL, Calkins EV.* Self-evaluation in undergraduate medical education: a longitudinal perspective. *J Med Educ* 1985;60:21-8.
10. *Rezler AG.* Self-assessment in problem-based groups. *Med Teacher* 1989;11:151-6.
11. *Risucci DA, Tortolani AJ, Ward RJ.* Ratings of surgical residents by self, supervisor and peers. *Surg Gyn Obs* 1989;169:519-26.
12. *Donald JB, Thomson D.* Evaluation of a computerized assessment package for general practitioner trainees. *Br J General Pr* 1993;43:115-8.
13. *Woolliscroft JO, Ten Haken J, Smith J, Calhoun JG.* Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med* 1993;68:285-94.
14. *Harrington JP, Murnaghan JJ, Regehr G.* Applying a relative ranking model to the self-assessment of extended performances. *Adv Health Sci Educ* 1997;2:17-25.
15. *Tracey JM, Arroll B, Richmond DE, Barham PM.* The validity of general practitioners' self assessment of knowledge: cross sectional study. *BMJ* 1997;315:1426-8.
16. *Johnson D, Cujec B.* Comparison of self, nurse, and physician assessment of residents rotating through an intensive care unit. *Crit Care Med* 1998;26:1811-6.
17. *Martin D, Regehr G, Hodges B, McNaughton N.* Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Acad Med* 1998;73:1201-6.
18. *Fox RA, Ingham Clark CL, Scotland AD, Dacre JE.* A study of pre-registration house officers' clinical skills. *Med Educ* 2000;34:1007-12.
19. *Friedlich M, MacRae H, Oandasan I, Tannenbaum D, Batty H, Reznick R, Regehr G.* Structured assessment of minor surgical skills (samss) for family medicine residents. *Acad Med* 2001;76:1241-6.

20. *Jansen JJM, Grol RPTM, Crebolder HFJM, Rethans JJ, van der Vleuten CPM.* Failure of feedback to enhance self-assessment skills of general practitioners. *Teach Learning Med* 1998;10:145-51.
21. *DiMatteo MR, DiNicola DD.* Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Med Care* 1981;19:829-39.
22. *Calhoun JG, Ten Haken JD, Woolliscroft JO.* Medical students' development of self- and peer-assessment skills: a longitudinal study. *Teach Learn Med* 1990;25-9.
23. *Kegel-Flom P.* Predicting supervisor, peer, and self ratings of intern performance. *J Med Educ* 1975;50:812-5.
24. *Speechly M, Weston WW, Dickie GL, Orr V.* Self-assessed competence: before and after residency. *Can Fam Phys* 1994;40:459-64.
25. *Marel GM, Lyon PM, Field MJ, Barnsley L, Hibbert E, Parise A.* Clinical skills in early postgraduate medical training: patterns of acquisition of confidence and experience among junior doctors in a university teaching hospital. *Med Educ* 2000;34:1013-5.
26. *Dubois V, Everwijn S, van Geldorp G et al.* The construction of a new curriculum of postgraduate training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
27. *Lamberts H, Wood M, editors.* *ICPC. International Classification of Primary Care.* Oxford: Oxford University Press; 1987.
28. *Bleys FC, Gerritsma JGM, Netjes I.* Skills development by medical students and the influence of prior experience: a study using evaluations by students and self-assessment. *Med Educ* 1986;20:234-9.
29. *Coutts L, Rogers J.* Student assessment and standardized patients-will the questions never end? *Acad Med* 1999;74(10Suppl):S128-30.
30. *Kramer AWM, Düsman H, Tan LHC, Jansen JJM, Grol RPTM, van der Vleuten CPM.* Effect of extension of postgraduate training in general practice on the acquisition of knowledge of trainees. *Fam Pract* 2003;20:207-12.
31. *Kramer AWM, Zuithoff P, Düsman H, Jansen JJM, Tan LHC, Grol RPTM, van der Vleuten CPM.* Predictive validity of a written knowledge test of skills in postgraduate training for general practice. *Med Educ* 2002;36:812-9.
32. *Cronbach LJ.* How we should measure 'change'- or should we? *Psychol Bull* 1970;74:68-80.
33. *Anonymous.* SPSS Advanced Statistics 7.5. Chicago: SPSS Inc; 1997: pp. 89-130.
34. *Kramer AWM, Jansen JJM, Düsman H, Tan LHC, van der Vleuten CPM, RPTM Grol.* Acquisition of clinical skills in postgraduate training for general practice. Accepted for publication in *Br J General Pr.*
35. *Kramer AWM, Düsman H, Tan LHC, Jansen JJM, Grol RPTM, van der Vleuten CPM.* Acquisition of communication skills in postgraduate training for general practice. Submitted to *Med Educ.*





GENERAL CONCLUSIONS AND DISCUSSION

Introduction

In many countries, among which the Netherlands, the general practitioner (GP) has a central role in the health care system. Consequently, postgraduate training in general practice should teach effectively trainees the knowledge, skills and attitudes that are required to perform the many and complex tasks of a GP.^{1,3} However, little is known about the effects of postgraduate training in general practice on the acquisition of the clinical competence of trainees.^{4,6} Evaluation of this acquisition was the aim of this thesis. We examined to what extent postgraduate training in the Netherlands achieves a satisfactory level of clinical competence at completion of training, and we evaluated the contribution of the training programme to the acquisition of clinical competence during the course. In order to achieve this, we formulated a set of specific research questions:

1. What is the acquisition of knowledge during a two-year and three-year postgraduate training programme for general practice, and does a three-year programme lead to a higher knowledge level than a two-year programme, while maintaining the same structure and learning objectives? (Chapter 1)
2. What are the credibility and reliability of a rational and an empirical standard setting procedure in an Objective Structured Clinical Examination (OSCE), and what is the impact of a reality check on the credibility and reliability of the rational procedure? (Chapter 2)
3. What is the predictive validity of a Knowledge Test of Skills for performance on an OSCE in postgraduate training for general practice? (Chapter 3)
4. What is the acquisition of clinical skills during a three-year postgraduate training programme for general practice, and is a satisfactory level of clinical skills achieved at completion of training? (Chapter 4)
5. What is the acquisition of communication skills during a three-year postgraduate training programme for general practice, and is a satisfactory level of communication skills achieved at completion of training? (Chapter 5)
6. What is the increase of self-perceived clinical competence of trainees in a three-year postgraduate training programme for general practice, and how can the growth of self-perceived clinical competence be explained? (Chapter 6)

In this chapter the main findings will be outlined and discussed, and recommendations for the future will be given.

Main findings

We will initially discuss the outcome of the two methodological studies. The first of these concerned standard setting for an OSCE (chapter 2). It was found that a newly developed empirical procedure related to the Borderline Group method, the Borderline Regression method, was more credible, reliable and feasible compared to the rational Angoff procedure. The latter was modified to obtain a feasible procedure; a group discussion and adjusted second judgment were omitted. This finding provides additional evidence that the Angoff procedure may be less suitable for performance-based skill testing than for written knowledge testing.⁷ Based on these results, the Borderline Regression method was chosen to judge the level of clinical skills at the end of the training (chapter 4). In the second study (chapter 3) it was shown that a written knowledge test of skills can be used as an instrument to estimate the level of clinical skills in postgraduate training for general practice. This especially applies for group evaluation, such as in studies examining the effects of a training programme or as a screening instrument for deciding about courses to be offered. However, it does not imply that written testing of knowledge of skills can replace the assessment of skills with an OSCE, as direct observation of performance of skills is more valid and stimulates trainees to practise them. Based on the results of this study, a knowledge test of skills was used to examine the acquisition of clinical skills throughout the course (chapter 4).

The first main issue of this thesis was to examine to what extent trainees, at completion of their training, have satisfactory clinical competence. The study on knowledge (chapter 1) showed that an extension of the training programme with one year led to a higher final knowledge level. The increase of knowledge obtained in both curricula was considerable, when compared to the increase of knowledge obtained in medical school.⁸ The results indicate that postgraduate training can contribute to the acquisition of knowledge to a sufficient extent in both curricula and that extension of the curriculum can further increase the level of knowledge. However, as we did not have a standard to judge whether the final level was satisfactory, firm conclusions cannot be drawn.



For clinical skills (chapter 4) trainees proved to have a satisfactory level at completion of a three-year training programme. Integral clinical skills, dealing with complete patient encounters, were somewhat better mastered than the more technical skills. The study on communication skills (chapter 5) revealed, on the other hand, an unsatisfactory level at the end of the same training programme. Finally, in the study on self-perceived clinical competence (chapter 6) we showed that trainees considered themselves sufficiently competent for knowledge, clinical skills and consultation skills at the end of three years training.

The second main issue was to examine to what extent clinical competence improved during the course. The training programme under study was characterised by a clearly outlined structure and curriculum design.⁹ It comprised three blocks of equal length, each with specified learning objectives and a specific learning environment. In the two-year curriculum each block took about eight months. With the start of the three-year curriculum, blocks of one year were introduced. The first and third blocks were reserved for practical training in a general practice, whereas the second block was dedicated to rotations through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. In addition to the practical training, one day per week was reserved for theoretical education, skills training and reflection, and a half-day per week was scheduled for self-directed learning. The results revealed that, with the exception of communication skills, the level of clinical competence at the end of training was higher than at the start. For knowledge and clinical skills, we assessed the increase at several intervals during the course (chapters 1 and 4) and we showed that the second block seemed to contribute less to the acquisition of knowledge and clinical skills relevant to general practice than the first and third block. As we did not find an increase in communication skills between start and finish of the three-year curriculum (chapter 5), it is perhaps possible that communication skills decreased in the second block. All this indicates the need to examine more closely the influence of the second block on the development of competence of trainees and on the effectiveness of the training programme as a whole. Finally, it was shown that the three-year curriculum enhanced self-perceived consultations skills more than self-perceived knowledge and clinical skills (chapter 6). This is contradictory to the results obtained by the objective assessment of clinical competence. Apparently, an increase in

objectively measured clinical competence does not explain an improvement of self-perceived clinical competence. In addition, no relation was found between the growth of self-perceived competence and several background variables that may potentially have an influence on it (chapter 6). As variables were recognised: growth of general knowledge and knowledge of skills, age, gender, prior medical experience, the effort someone has spent upon her/his education and insight in weak and strong areas of clinical competence. So, although a three-year training programme improves self-perceived clinical competence, we still do not know to what it refers and by what it is enhanced. As self-assessment of clinical competence is widely used in medical education and the importance of it for self-directed learning and professional autonomy is well established,¹⁰⁻¹¹ further study into the theoretical concept of self-assessment and into the factors contributing to the feeling of being competent, is required.

Finally, in two studies (chapters 4 and 5) we compared the results of trainees to those of experienced GPs. Because practical learning in a general practice comprised the greatest part of the curriculum, the experienced GP had an important model role for trainees. Our studies revealed that trainees, at completion of their training, performed better than (clinical skills) or comparable to (communication skills) experienced GPs. The consequences of this finding for the teaching and learning of clinical and communication skills and for the model function of the GP-trainer, need further investigation.



Methodological considerations

In this thesis the naturalistic environment of postgraduate training was investigated to examine the influence of the training programme on the acquisition of clinical competence during postgraduate training.¹² The benefit of this approach is that the results provide evidence from the real world rather than from an artificial, controlled situation, which implies a greater relevance.¹³ However, since a naturalistic environment is subjected to many factors that are beyond the control of the investigators, interpretation of the results is more difficult. Furthermore, as the intervention (the entire training programme) was complex, it is also hard to distinguish which part of the training programme is responsible for which effect. Given these limitations, we have aimed at the best available evidence by examining the change of knowledge and skills across years of training.¹⁴

For the comparison of the two- and three-year curriculum (chapter 1) we collected over the years a large number of knowledge test scores. This mixed longitudinal design is considered an optimal design for the measurement of change over time.¹⁵ For the acquisition of clinical skills (chapter 4) we assessed the knowledge of skills in a cross-sectional design. Finally, for the acquisition of communication skills (chapter 5) and the development of self-perceived clinical competence (chapter 6) we applied a longitudinal design with a start and end measurement. Because this study covered a period of many years and was performed in ongoing curricula, participation was more influenced by the course of events and less controlled by the investigators compared to an experimental study. When possible the effect of this unintended selection was evaluated by analysing differences between participating and missing trainees. These analyses revealed no statistically significant differences. For the acquisition of general knowledge and knowledge of skills this procedure could not be applied, because data to analyse differences were not available. However, as large groups of trainees participated in these studies, selection effects are less likely to occur. In most studies the sample size was acceptable (from 80 to 800 participants on average). However, in two studies (chapters 3 and 6) it was rather small (47 and 25 participants). Although this size limits general conclusions, the fact that participants in both studies were representative for the entire group, according to several criteria such as age, gender, time between graduation and start postgraduate training, entrance level of knowledge and knowledge of skills, is reassuring.

Discussion and recommendations

National assessment

A national evaluation was the subject of this investigation. Such national evaluation was made possible, because the eight training institutes had accepted the guidelines for the curriculum, as proposed by the committee 'Construction of a new Curriculum of Postgraduate Training for General Practice' (CCBOH),⁹ and as a consequence their training programmes and learning objectives provided a sufficient degree of uniformity.¹⁶ However, it is very likely that on a detailed level differences remained between the institutes, for instance in the way they administered the training programme, and these differences could possibly influence the acquisition of clinical competence of trainees. As this study was part of a comprehensive evaluation, differences

between the institutes have been addressed as well.¹⁶ The results provide evidence that institutes differed in the extent to which the knowledge and clinical skills of their trainees increased during training.^{16,17} This is interesting information, because it may reveal more insight in the relation between the process of education and the outcome. Therefore, a closer examination of differences between institutes is recommended (see also next paragraph).

Process evaluation

This thesis focused on the outcome of postgraduate training. Although outcome measures are more directly related to the quality of the training programme, evaluation of the process of education is necessary and valuable as well.^{18,19} In the introduction chapter of this thesis we have described the structure, design and learning objectives of the curriculum. According to the model of Remmen this is referred to as 'the intended curriculum'.²⁰ What matters is how the intended curriculum has been implemented ('the process' according to Donabedian²¹ or 'the curriculum in action' according to Remmen²⁰). Insight in this part of the curriculum provides a more complete impression of the educational activities and enables connections between the different aspects of education, i.e. the intended curriculum, the curriculum in action (process) and the learned curriculum (outcome). This especially applies for Dutch postgraduate training in general practice, since a national intended curriculum is supposed to be administered by eight different training institutes. Important issues are how the national guidelines are interpreted by each institute; whether an institute has additional mission statements and learning objectives; which teaching and assessment procedures it uses; how it is organised etc. All these aspects may influence the outcome of the educational activities.

In the comprehensive evaluation, which this study was part of, the curriculum in action has been subject of four studies.¹⁶ The results of these studies are briefly outlined in Box 1.

The results concerning the second year (see the third study in Box 1) may explain the findings in this thesis, that the second year seems to contribute less to the acquisition of clinical competence than the first and third year. From the literature into undergraduate medical education during clinical rotations, we know that this education may be problematic.²³ Medical teams are often unable to devote sufficient time and effort to teaching.



Moreover, the learning environment is subjected to many uncontrollable factors, such as patient mix and urgent cases. And finally, medical teams have often difficulty to specify learning objectives. This probably also applies for clinical rotations in postgraduate training. It emphasises again the need for further research of this part of the curriculum both into the educational process and the outcome.

Of interest is also the trainees' appreciation of the education and the discrepancy between their self-perceived clinical competence and the outcome of the objective assessment of this competence. The observation that their self-perceived consultation skills increased more than their self-perceived clinical skills, is congruous with their appreciation of the training in consultation skills. It is perhaps possible that trainees thought they were competent in consultations skills due to good consultation skills training and therefore studied consultations skills less resulting in low objectively measured scores of communications skills. Of course this very hypothetical explanation needs to be further investigated, but it emphasises an important issue as advocated by ten Cate: the effect of educational activities on the learning behaviour of learners.²⁴



Assessing clinical competence

Our approach to assessing clinical competence can be characterised as follows. Firstly, we have applied the conceptual framework of Miller.²⁵ In this model assessment is approached as a pyramid with knowledge at the base, subsequently followed by 'knows how' and 'shows how', and with 'does' at the top. Secondly, we have used assessment instruments in which competence is measured by quantitative, explicit, and, as far as possible, objective criteria (e.g. knowledge tests, checklists). This approach has some benefits. Subdivision of the several aspects of clinical competence enhances specified and detailed feedback, which facilitates the learning process of trainees. Furthermore, there is sufficient evidence that knowledge is core to any clinical competence,²⁶ indicating its relevancy. And, finally, quantitative, explicit and objective assessment criteria guarantee a more standardised and comparable judgement of clinical competence.²⁷ However, there are also some drawbacks. Our assessment instruments have difficulty in measuring the professional aspects of clinical competence. In the general practice context examples of these are: the logic and suppleness with which the patient

Box 1

Summary of the four studies into 'the curriculum in action'

The **first** study included the educational encounters between trainee and GP-trainer. It was investigated to what extent the guidelines of the committee CCBOH for these encounters were applied and how the trainees judged their quality. The results showed that the guidelines were not completely applied: instead of one encounter each day, trainees received education once in two days; the content and educational approach of the encounters in the first and third year of training did not differ although the committee CCBOH proposed the encounters to be guided by the specific learning objectives of these years; and there was no attunement to the education at the institute. In general, the trainees appreciated the quality of the encounters. The quality would improve by a better preparation of the encounter, by more follow-up activities and by more positive feedback.²²

In the **second** study it was examined how often institutional education was offered; what the content of the programme was; and whether differences existed between and within training institutes. It was found that, on average, 50 days a year institutional education was organised. This number is sufficient according to the guidelines of the committee CCBOH. The content of the programme included for 70% the four main tasks of the GP. Task 2, tasks related to certain categories of patients, complaints and diseases, received the greatest attention followed by task 1, tasks related to the process of providing care, and task 4, tasks related to the doctor as a person. Differences in content of the programme were found between and within training institutes.

The **third** study concerned the second year, in which trainees rotated through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. Research questions were how the education in this period was performed and how the trainees judged its quality. The results showed that, on average, the hospital clerkships took six months and the clerkships in clinics for chronically ill patients and psychiatric outpatients clinics each three months. The majority of the hospital clerkships included the specialisations 'Surgery' and 'Internal Medicine', followed by 'Psychiatry' and 'Gynaecology'. The specialisations 'ENT' and 'Ophthalmology' took each only one percent of all clerkships. The trainees generally judged their activities during the clerkships as relevant, but found that the hospital clerkships included less relevant activities such as routine tasks and surgical assistance. They experienced difficulties in the learning plan and objectives, the self-directed learning opportunities, the extent to which the practical education was attuned to the institutional education and the freedom they had to choose specialities of their own interest or learning needs.

Finally, in the **fourth** study trainees were asked to give their opinions about the quality of each training year and the training programme as a whole. They highly appreciated the learning environment of the general practice, whereas the learning during the clinical rotations and at the training institutes was somewhat less valued. They estimated that they learned most in the first year followed by the third and then the second year. And finally, they appreciated across the three years the education in knowledge relevant to general practice and consultation skills more than the education in clinical skills.



problem is approached, the extent to which continuity of care is applied and the more experienced-based aspects of care. These aspects are especially of interest in the postgraduate setting, in which trainees already have experience in and are prepared to practising as a doctor.²⁸ Apart from this, our instruments have difficulty in assessing attitude, an important aspect of clinical competence as well. Moreover, since clinical competence is a complex phenomenon, which requires a combination and integration of knowledge, skills and attitude, mastering of individual aspects of competence, as is assessed by us, does not necessarily mean that the doctor will be able to integrate them.^{29,30}

From these arguments it will be clear that the assessment of these more professional aspects of clinical competence requires other tools than the ones applied in our study. It appears that more qualitative measures, which rely on professional and patient judgement, can play a useful role in this.^{29,31-33} Since professional and patient judgement implies a subjective judgement, the concern is the reliability and validity. Therefore, it is recommended that the judgements be made on a number of occasions and by different persons.³³



A broader approach to clinical competence

Aim of our study was to evaluate to what extent trainees, at completion of their training, master the tasks of the GP as defined by the Basic Job Description.³⁴ However, due to feasibility reasons, only two of the four main tasks have been addressed: task 1, tasks related to the process of providing care and task 2, tasks related to certain categories of patients, complaints and diseases.⁹ These tasks refer to the more direct and traditional aspects of clinical competence in general practice. The two other tasks are related to the conditions under which the care is provided (task 3) and to the doctor as a person (task 4). These tasks comprise competencies concerning the maintenance of professional competence, the cooperation with other care providers, the organisation and management of a general practice, the social position of the GP, the personal view on the work of a GP, the influence of personal characteristics on the work and the attunement of the private to the professional life. All these tasks refer to a broader approach of competence in general practice.

In recent years the importance of broader outcomes of medical education has been emphasised.^{31,32,35,36} Harden et al. described a three-circle model representing the three main outcomes of medical

education: what the doctor is able to do, how the doctor approaches practice and the doctor as a professional.³⁶ Murray et al. recognized several dimensions of potential outcomes of medical education including, doctors and patients attributes.³² Examples of these are professional behaviour, efficiency, team working, patient satisfaction and improved health of patients. They also examined to what extent available assessment procedures could measure these outcomes. Their conclusion was that extensive tools for assessing knowledge and clinical skills existed, but that more work was to be done in assessing the broader range of competencies.

If postgraduate training in general practice wants to demonstrate that trainees possess these broader aspects of clinical competence, it has to develop and implement reliable and valid instruments for assessing them.

Gender

In our studies we did not consider gender differences. From the literature we know that learning styles of men and women differ.³⁷ The areas of interest may vary as well. For instance, men are probably more fascinated by technical procedures, while women are perhaps more concerned about communication aspects.³⁸ Also the gender of the trainee and the GP-trainer may have effect on the learning process of the trainee, because it influences the role model function and identification process. Another aspect is the impact of gender on the rating of performance. There is evidence that female observers judge the performance of trainees different from male observers and that the gender of the trainees has influence on the judgement.³⁹ Both aspects are important issues to be considered in the future, when evaluating the effects of educational activities.

Approach to research

We have adopted a (quasi-) experimental research approach. Such approach includes hypothesis formulation and quantitative data gathering and analysis with focus on prediction and generalisation. Although this approach yields interesting and useful findings and may provide guidelines for teaching practices, it also has its limitations.⁴⁰⁻⁴³ Examples of these are: it only studies what can be quantitatively measured; it may be biased, due to underlying assumptions of the investigator that may prevent renewing ideas; it contains uncontrolled factors, especially in the field of medical education.

More generally, experimental, quantitative research can help to answer questions like 'how many' and 'how often'. Because learning and teaching involve many more dimensions, emotional, cognitive, attitudinal and physical, and because the process of learning is personal, subjective and highly variable, a qualitative research approach is required as well that may shed light on the questions of 'why' and 'how'.^{13,40-42} For instance, how is the interaction between teachers/trainers and students, how do students perceive educational activities or what is the effect of a teaching procedure on their learning behaviour? To gain a better understanding of medical education, both quantitative and qualitative research approaches are required. Depending the problem at hand, the investigator may choose the appropriate approach. Often both approaches will complement each other, as is shown in our study in which several issues have emerged that need further qualitative examination. Examples of these are: self-assessment and self-perception of clinical competence, the learning of communication skills, learning styles and interests of female and male trainees, and the interaction between GP-trainer and trainee.



Practical recommendations for the near future

This thesis has shown that national evaluation of postgraduate training in general practice is feasible. Two factors have greatly contributed to the successful realisation of this study: the existence of a national centre for evaluation that closely cooperates with the eight training institutes; and the use of assessment instruments that are developed and implemented by this centre, except for the MAAS-Global list. To continue such studies, a national centre for evaluation, bringing together qualified educationalists and GPs, remains an essential condition. Ideally speaking, this centre has to be professionally independent but intimately linked with the eight institutes. Apart from evaluation studies, an important task of the centre should be the development and validation of new instruments for the evaluation of the structure, process and outcome of postgraduate training, to improve the quality of evaluation. It is recommended that means will be provided to enable a continuous evolution of evaluation tools and that the expertise of medical educationalists will be taken into consideration in this renewal process.⁴⁴ Another area of interest for this centre would be the development of the educational process itself; for instance the innovation of the curriculum, the professionalisation of


teachers and trainers and the teaching of communication skills. As process and outcome are mutually interdependent, compartmentalising of both aspects is unproductive and not advisable.⁴⁴ Moreover, it is more economical to involve both aspects, since the educational expertise of the centre can be applied to a greater extent.

In the next years, the curriculum will change due to newly formulated learning outcomes and the provision of a new blueprint. The main differences with the 'old' blueprint are that the teaching is organised according to the principles of modular teaching and self-directed learning, and that trainees may receive dispensation and/or may choose a GP-specialisation.^{45,46} Evaluation will provide the necessary evidence that the objectives of the new curriculum are indeed achieved. From the findings presented in this thesis, the following recommendations can be used for the near future.

In the curriculum examined in this study, hospitals, clinics for chronically ill patients and psychiatric outpatients clinics were chosen as a practical learning environment for the acquisition of additional knowledge, skills and attitude that cannot be learned in a general practice. It was also decided that this education has to be placed between two periods of practical learning in a general practice, to enable that it was perceived from the learning perspective of a GP-trainee. As our findings indicated that trainees acquired less knowledge and skills relevant to general practice in this period, we question (1) to what extent the chosen learning environment and its position in the curriculum are appropriate to achieve the learning objectives of this period and (2) to what extent our assessment instruments were capable to measure the knowledge and skills acquired during this period. For the future it will therefore be essential to redefine which knowledge, skills and attitude are necessary for a GP that cannot be learned in the learning environment of a general practice; which learning environment will be appropriate for this purpose; when and how this education has to take place; and how it can be assessed. Special issues of interest with these are the consequences of this education for the other learning objectives and the suitability of hospitals for this purpose, since they are increasingly becoming centres where patients only stay for a short time, to undergo highly specialised investigation or treatment. Moreover, considering the predicted shortage of GPs, it has to be taken into consideration

whether a shortening of the curriculum is acceptable without causing shortcomings in the quality of future GPs. Perhaps the training time devoted to learning outside a general practice can be reduced. Another solution could be that the last year of medical school becomes a part of postgraduate training in general practice, as is already the case in Belgium.⁴⁷ The learning objectives that cannot be achieved in a general practice could be learned then. This solution corresponds to the discussion about the medical education continuum in the Netherlands that recently has resulted in recommendations for a new approach to medical education from medical school to continuous professional development.⁴⁸

As adequate communication skills are crucial for good clinical practice, the acquisition of these skills requires full attention. The results of our study into communication skills are material for new questions to be answered: what are essential communication skills for general practice, how can they be learned and put into practice, what is the influence of the several parts of the curriculum on the acquisition of these skills and how can they be assessed?



Since practical learning in a general practice comprises the greatest part of the training programme, the role of the experienced GP as a learning model for trainees is of great influence. From our findings we know that their communication and clinical skills are not better than those of the trainees. What does that mean for the learning of these skills? Are there other skills in which GP-trainers are more competent than trainees? Does a GP-trainer have to be more competent than a trainee or is her/his role first of all to support the learning of the trainee? What are important teaching skills for GP-trainers and how can they be learned and assessed? The last question has recently been addressed by Boendermaker et al.^{49,50} Thirty-seven characteristics of the competent GP-trainer were identified, of which the provision of feedback, the ability to communicate with the GP-trainee, respect for the GP-trainee and the ability to inspire reflection in the GP-trainee were judged as most important. The validity and reliability of an Objective Structured Teaching Examination (OSTE), a self-assessment questionnaire and a GP-trainee assessment questionnaire for evaluating the educational competence of GP-trainers were investigated. It was found that more research into the assessment of the educational competence of GP-trainers is required and that a variety of assessment procedures are needed to cover its complexity. Finally, from the literature we know that modelling is a powerful force

in the education of future doctors.¹⁰ How does a GP-trainer influence the professional behaviour of the trainee? These are all important issues that need further attention.

An important goal of educational evaluation is to examine whether learning outcomes are achieved. For this purpose it can be necessary to set a standard for adequacy. We have examined standard setting procedures for an OSCE. Comparable studies are necessary for the knowledge tests. Since standard setting is a judgemental process, the credibility and reliability are important subjects of investigation. But also the feasibility of the procedure has to be considered, to guarantee its use.

Finally, it will be interesting to learn more about self-assessment of clinical competence. To which aspects of clinical competence and/or personal characteristics does it refer? How does it influence self-directed learning and professional autonomy? Which aspects of the training programme contributes to an increase of self-perceived competence? What is the effect of it on the learning behaviour? Since a theoretical concept of self-assessment in (continuing) medical education is not available, it would be worthwhile to develop such a concept. This requires a profound examination of this important aspect of educational and professional evaluation.

In conclusion, this study has provided ample evidence of the usefulness of evaluation of medical education, in particular in the field of postgraduate training in general practice. Our belief is that this is just a start of a much broader evaluation of postgraduate training, an event necessary for the development of a highly competent GP. May this study be a catalyst for a critical and more evidence-based approach to the teaching of future GPs.

References

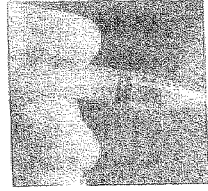
1. *Tarimo E, Webster EG.* Primary health care: concepts and challenges in a changing world. Alma-Ata Revisited. 1-118. Geneva: WHO; 1997.
2. *Haq C, Ventres W, Hunt V, Mull D, Thompson R, Rivo M, Johnson P.* Where there is no family doctor: the development of family practice around the world. *Acad Med* 1995;70:370-80.
3. *Prystowsky JB, Bordage G.* An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ* 2001;35:331-6.
4. *Spike NA, Hays RB.* Analysis by training status of performance in the certification examination for Australian family doctors. *Med Educ* 1999;33:612-5.
5. *Van Leeuwen YD.* Growth in knowledge of trainees in general practice [dissertation]. Maastricht: Universitaire Pers Maastricht; 1995.
6. *Irby D.* Teaching and learning in ambulatory care setting: a thematic review of the literature. *Acad Med* 1995;70:898-931.
7. *Kane MT, Crooks TJ, Cohen AS.* Designing and evaluating standard setting procedures for licensure and certification tests. *Adv Health Sci Educ* 1994;4:195-207.
8. *Verhoeven BH, Verwijnen GM, Scherpbier AJJA, van der Vleuten CPM.* Growth of medical knowledge. *Med Educ* 2002;36:711-7.
9. *Dubois V, Everwijn S, van Geldorp G et al.* The construction of a new curriculum of vocational training for general practice in the Netherlands. Utrecht: Royal Dutch Medical Association (KNMG); 1987.
10. *Irvine D.* The performance of doctors. I: Professionalism and self regulation in a changing world. *BMJ* 1997;314: 1540-2.
11. *Oermann MH.* Professional reflection: have you looked in the mirror lately? *Orthopaedic Nursing*;1998:22-6.
12. *Dimitroff A, Davis WK.* Content analysis of research in undergraduate medical education. *Acad Med* 1996;71:60-7.
13. *Buckley G.* Partial truths - research papers in medical education. *Med Educ* 1998;32:1-2.
14. *Harden RM, Grant J, Buckley G, Hart IR.* BEME Guide No.1: Best Evidence Medical Education. *Med Teacher* 1999;21:553-62.
15. *Van 't Hof MA, Roede MJ, Kowalski CJ.* A mixed longitudinal data analysis model. *Hum Biol* 1977;49:165-179.
16. *Tan LHC, Kramer AWM.* Opleiding doorgelicht (National evaluation study of Dutch postgraduate training in general practice). Utrecht: SVUH; 1999.
17. *Tan LHC, Düsman H, Kramer AWM, Jansen JJM.* Evaluatie driejarige huisartsopleiding. EVA-H. Resultaten per afdeling (National evaluation study of Dutch postgraduate training in general practice. Results of the individual training institutes). Utrecht: SVUH; 2000.
18. *Elliott RL, Juthani NV, Rubin EH, Greenfeld D, Skelton WD, Yudkowsky R.* Quality in residency training: toward a broader, multidimensional definition. *Acad Med* 1996;71:243-7.

19. *Yudkowsky R, Elliott RL, Schwartz A.* Two perspectives on the indicators of quality in psychiatry residencies: program directors' and residents'. *Acad Med* 2002;77:57-64.
20. *Remmen R.* An evaluation of clinical skills training at the medical school of the University of Antwerp [dissertation]. Antwerpen: Cuypers; 1999.
21. *Donabian A.* The quality of care. How can it be assessed? *JAMA* 1988;260:1743-8.
22. *Boendermaker PM, Ket P, Düsman H, Schuling J, van der Vleuten CPM, Tan LHC.* What influences the quality of educational encounters between trainer and trainee in vocational training for general practice? *Med Teacher* 2002;24:540-3.
23. *Wolffhagen HAP, Gijssels WH, Dolmans D, Essed G, Schmidt HG.* Improving clinical education through evaluation. *Med Teacher* 1997;19:99-103.
24. *Ten Cate O.* What happens to the student? The neglected variable in educational outcome research. *Adv Health Sci Educ* 2001;6:81-8.
25. *Miller GE.* The assessment of clinical skills/competence/performance. *Acad Med* 1990;65 (9Suppl):S63-7.
26. *Schmidt H, Norman G, Boshuizen H.* A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990;65:611-21.
27. *Van der Vleuten CPM, Norman GR, de Graaff E.* Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ* 1991;25:110-8.
28. *Norman GR, van der Vleuten CPM, de Graaff E.* Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ* 1991;25:119-26.
29. *Hager P, Gonzi A, Athanasou J.* General issues about assessment of competence. *Assessment & Evaluation in Higher Education* 1994;19:3-16.
30. *Harden RM.* Learning outcomes and instructional objectives: is there a difference? *Med Teacher* 2002;24:151-5.
31. *Gray JD.* Global rating scales in residency education. *Acad Med* 1996;71(1Suppl):S55-63.
32. *Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO.* The accountability of clinical education: its definition and assessment. *Med Educ* 2000;34:871-9.
33. *Van der Vleuten C.* Validity of final examinations in undergraduate medical training. *BMJ*;321:1217-9.
34. *Springer MP, editor.* Basic Job Description for the general practitioner. Utrecht: Dutch National Association of General Practitioners (LHV); 1983.
35. *Harden RM, Crosby JR, Davis MH.* AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. *Med Teacher* 1999;21:7-14.
36. *Harden RM, Crosby JR, Davis MH, Friedman M.* AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. *Med Teacher* 1999;21:546-52.
37. *De Saintonge DM, Dunn GM.* Gender and achievement in clinical medical students: a path analysis. *Med Educ* 2001;35:1024-33.
38. *Van den Brink-Muinen A.* Gender, health and health care in general practice. Utrecht, Nivel; 1996.



39. *Rutala PJ, Witzke DB, Leko EO, Fulginiti JV.* The influences of student and standardized patient genders on scoring in an objective structured clinical examination. *Acad Med* 1991;66(9Suppl): S28-30.
40. *Coles CR, Grant JG.* Curriculum evaluation in medical and health-care education. *Med Educ* 1985;19:405-22.
41. *Harden RM.* Approaches to research in medical education. *Med Educ* 1986;20:522-31.
42. *Irby DM.* Shifting paradigms of research in medical education. *Acad Med* 1990;65:622-3.
43. *Vroeijenstijn AI.* Quality assurance in medical education. *Acad Med* 1995;70(7Suppl):S59-67.
44. *Van der Vleuten C, Scherpbier A.* De rol van de onderwijskundige in de beroepsopleiding (The role of the educationalist in postgraduate medical training). In: *Kramer A, Jansen J, redactie. Waardering van onderwijs (Valuation of education).* Utrecht, SVUH; 2000.
45. *Wigersma L, van Berkesteijn LGM, Giesen P.* Eindtermen Huisartsopleiding 2000 (Learning outcomes Postgraduate Training in General Practice 2000).
46. *Onstenk JHAM, Swager R.* Raamplan Huisartsopleiding (Blueprint postgraduate training in general practice). 's Hertogenbosch: Cinop; 2001.
47. *Denekens J.* Kwaliteitszorg voor de huisartsopleiding in Vlaanderen (Quality assurance of postgraduate training in general practice in Flanders). In: *Kramer A, Jansen J, redactie. Waardering van onderwijs (Valuation of education).* Utrecht, SVUH; 2000.
48. www.knmg.nl/publicaties: Arts van straks; een nieuw medisch opleidingscontinuüm (The future physician; a new medical education continuum); 2002.
49. *Boendermaker PM.* Meesterschap. Van verkenning naar herkenning van de goede huisartsopleider (Mastership. From exploration to recognition of the competent GP-trainer) [dissertation]. To be published (April 2003).
50. *Boendermaker PM, Schuling J, Meyboom-de Jong B, Zwierstra RP, Metz JCM.* What are the characteristics of the competent general practice trainer? A qualitative study examining the characteristics of the competent GP-trainer. *Fam Pract* 2000;17:547-53.





SUMMARY OF THIS THESIS

In this thesis we evaluate the development of clinical competence during postgraduate training in general practice. After an introduction of the topic, six chapters deal with the results obtained by our study and this is followed by general conclusions and a discussion.

Introduction

Due to an increasing demand and growing complexity of the provision of primary health care, several countries have, over the last 30 years, introduced postgraduate training programmes for general practice. Although all these programmes aim to educate trainees adequately, little evidence exists whether their programmes result in qualified general practitioners (GPs). Therefore, evaluation studies are required. This thesis deals with the evaluation of Dutch postgraduate training in general practice.

The development of this training finds its origin in 1973, when the eight training institutes, attached to the departments of general practice of the eight medical schools, started its provision. Initially, the training took one year. In 1985, a committee was formed that took care of the 'Construction of a new Curriculum of Postgraduate Training for General Practice' (CCBOH). This committee has provided national guidelines for the structure, duration and learning objectives of the curriculum. Basis for the content was the Basic Job Description of the GP formulated by the Dutch National Association of GPs (LHV). This job description recognised four main tasks for the GP: (1) tasks related to the process of providing care in general practice; (2) tasks related to certain categories of patients, complaints and diseases; (3) supportive tasks: ability to organise and manage a general practice, to cooperate with other care providers, and to maintain professional competence; and (4) tasks related to the personal functioning as a GP. The eight training institutes accepted these guidelines and this provided, from 1988 onwards, a greater degree of uniformity between their training programmes.

Although the committee CCBOH proposed a three-year curriculum, the government consented to a two-year curriculum in 1988. Because still shortcomings in the content were demonstrated, extension to three years, while maintaining the structure and learning objectives, was permitted in 1994.

Since 1987 the eight training institutes cooperate in a National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).

This centre performed a comprehensive national evaluation (EVA-H project) from 1995 to 1999 and this thesis governs part of that evaluation.

The evaluation presented in this thesis focused on the outcome and comprises the following questions: (1) to what extent do trainees, at completion of their training, master the tasks of the GP as defined by the Basic Job Description? (2) is there any difference between the two-year and the three-year curriculum? and (3) what is the acquisition of clinical competence during the course? The study was limited to the evaluation of the learning objectives related to the domains knowledge and skills of task 1, tasks related to the process of providing care, and task 2, tasks related to certain categories of patients, complaints and diseases. In addition, it concerns two methodological studies, in order to facilitate answering of the main research questions.

The specific research questions posed were:

1. What is the acquisition of knowledge during a two-year and three-year postgraduate training programme for general practice, and does a three-year programme lead to a higher knowledge level than a two-year programme, while maintaining the same structure and learning objectives? (Chapter 1)
2. What is the credibility and reliability of a rational and an empirical standard setting procedure in an Objective Structured Clinical Examination (OSCE), and what is the impact of a reality check on the credibility and reliability of the rational procedure? (Chapter 2)
3. What is the predictive validity of a Knowledge Test of Skills for performance on an OSCE in postgraduate training for general practice? (Chapter 3)
4. What is the acquisition of clinical skills during a three-year postgraduate training programme for general practice, and is a satisfactory level achieved at completion of training? (Chapter 4)
5. What is the acquisition of communication skills during a three-year postgraduate training programme for general practice, and is a satisfactory level achieved at completion of training? (Chapter 5)
6. What is the increase of self-perceived clinical competence of trainees in a three-year postgraduate training programme for general practice, and how can the growth of self-perceived clinical competence be explained? (Chapter 6)

Chapter 1

This chapter questions the assumption that a three-year curriculum would lead to a higher knowledge level than a two-year curriculum. Moreover, it examines to what extent the training programme contributed to the acquisition of knowledge throughout the two- and three-year course.

Knowledge was assessed by written progress-testing of knowledge relevant to general practice and embedded in real life situations.

Progress-testing implies that each test has the intention to reflect the cognitive final objectives of the curriculum. In a retrospective, mixed longitudinal design, test results of trainees who received the two-year curriculum (1992-1996) and the three-year curriculum (1994-1999) were compared.

The findings of the study indicate that the extension of the two-year curriculum with one year, while maintaining the same structure and learning objectives, leads to a higher knowledge level. The increase in both curricula was considerable when compared to the increase of knowledge in medical school. In the two-year as well as in the three-year curriculum most acquisition of knowledge occurred in the first eight months of training and towards the end of the course. Both curricula showed the largest stagnation in growth at the end of the second block, in which trainees rotated through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics.

As knowledge is core to any kind of clinical competence, this finding questions the impact of the second block on the development of competence of trainees and on the effectiveness of the training programme as a whole. Further research into these aspects is needed before more firm conclusions are to be drawn.

Chapter 2

The study described in this chapter is a methodological study concerning standard setting procedures. This study was necessary, to enable us to answer the second part of research question 4 (whether a satisfactory level of clinical skills was achieved at completion of training (see chapter 4). In order to decide whether a satisfactory level is achieved, tests results have to be compared to a standard of adequacy. Standard setting is a rather judgmental process, involving an arbitrary decision of what is considered to be 'good enough'.

To guarantee that this process is defensible and controllable, aspects of credibility and reliability have to be taken into account. This study focused on two absolute standard setting procedures for an OSCE. A modified Angoff method (written instructions only, without opportunity for discussion and an adjusted second judgement) with and without reality check was used as rational procedure. A new method related to the Borderline Group procedure, the Borderline Regression (BR) method, was used as empirical procedure. The reliability, credibility and the impact of a reality check of both procedures were examined. The OSCE of the study was administered to a national sample of trainees in their final (third) year of postgraduate training and a reference group of experienced GPs. Reliability was assessed using generalisability theory. Credibility was judged by comparing pass rates of trainees and experienced GPs, and by relating the passing scores to test difficulty.

The passing score was 73.4% for the Angoff method without reality check (Angoff I), 66.0% for the Angoff method with reality check (Angoff II) and 57.6% for the BR method. The reliabilities (expressed as root mean square errors) were 2.1% for Angoffs I and II, and 0.6% for the BR method. The pass rates of the trainees and GPs were 19% and 9% for Angoff I, 66% and 46% for Angoff II, and 95% and 80% for the BR method, respectively. The correlation between test difficulty and passing score was 0.69 for Angoff I, 0.88 for Angoff II and 0.86 for the BR method, respectively.

Based on these results, the BR procedure seems a more credible and reliable procedure to set a standard for an OSCE compared to the modified Angoff procedure, even when the latter is applied with a reality check. The overall conclusion is that the BR method provides a reasonable, defensible and feasible standard for an OSCE.

Chapter 3

The study described in this chapter concerns the second methodological study that enables us to answer the first part of research question 4 (see chapter 4). This time we focused on the assessment of clinical skills. The OSCE has been identified as one of the most valid methods for the assessment of these skills. However, since large-scale application of an OSCE is complex and expensive, alternative instruments with less demanding characteristics may be useful. We therefore studied the predictive validity of a written

knowledge test of clinical skills for performance on an OSCE in postgraduate training for general practice. The OSCE included technical, procedural skills and clinical skills related to complete patient encounters. Three hypotheses were tested referring to (1) the content of the knowledge test (general knowledge vs. knowledge of skills); (2) the rating system of the OSCE (global vs. checklist); and (3) the content of the OSCE (technical, procedural skills vs. complete patient encounters). A national sample of trainees of the three-year postgraduate training was subjected to a knowledge test of skills, a general knowledge test and an OSCE. Correlations between the various test results were analysed.

Two hypotheses were confirmed. Firstly, the knowledge test of skills correlated better to the OSCE than the general knowledge test, in particular for the OSCE stations including technical, procedural skills only (hypothesis 1). Secondly, the knowledge test of skills correlated better to the OSCE stations including technical, procedural skills than to the OSCE including complete patient encounters (hypothesis 3). Finally, the knowledge test of skills correlated better to the checklist rated than to the global rated OSCE stations including complete patient encounters (hypothesis 2). However, we found no important difference between the checklist rated and global rated OSCE stations including technical, procedural skills.

The results support the predictive validity of a knowledge test of skills. This implies that, in postgraduate training for general practice, a written knowledge test of skills can be used as an instrument to estimate the level of clinical skills, especially for group evaluation, such as in studies examining the effects of a training programme or as a screening instrument for deciding about courses to be offered. This estimation is more accurate when the content of the test matches the skills under study. However, the results do not mean that written testing of knowledge of skills can replace the assessment of skills with an OSCE, since direct observation of performance of skills is more valid and stimulates trainees to practise them.

Chapter 4

As clinical skills (including history taking, physical examination, diagnosing and management) constitute a crucial part of the daily work of GPs, it is one of the main tasks of postgraduate training in general practice to contribute to a satisfactory acquisition of these skills

by trainees. Therefore the study presented in this chapter examined the acquisition of clinical skills throughout the years of training as well as whether a satisfactory level was achieved at completion of training.

The postgraduate training of the study took three years and emphasised the acquisition of clinical skills. In a cross-sectional design the level of clinical skills throughout the years was measured, using a knowledge test of skills with a progress-testing format. The test was administered to trainees in all three years of training. For the assessment of the final level of clinical skills, a national sample of trainees, in their final phase of training, took both the knowledge test of skills and an OSCE. The OSCE included technical, procedural skills and clinical skills related to complete patient encounters (the so called integral stations). The results of both tests were compared to a standard of adequacy and to the results of a reference group of GP-trainers.

The increase of clinical skills was 18% (from 30% to 48%). The greatest acquisition took place in the first six months of training. Thereafter, a rather continuous but less rapid increase is observed, except for training level five in which the increase stagnated. At this level trainees completed their second year in which they rotated through hospitals, clinics for chronically ill patients and psychiatric outpatients clinics. The trainees performed better than the GP-trainers, both at the knowledge test and the OSCE. About 90% of them would have passed the knowledge test and 95% of them the OSCE. The results of the OSCE showed a remarkable distinction between the technical and integral stations. More trainees would have passed the integral stations than the technical stations, indicating that integral skills were better mastered than technical skills. Furthermore, the results of the integral stations were more or less similar for the trainees and GPs, while for the technical stations the GPs performed less than the trainees.

These findings indicate that a three-year postgraduate training programme for general practice, in which the acquisition of clinical skills is an essential goal, contributes to an increase of these skills. Integral skills dealing with complete patient encounters seem to be better acquired than the technical, procedural skills. The findings also suggest that the second year of the curriculum, in which trainees rotate through hospitals and the like, contributes less to the acquisition of clinical skills than the first and third years of training in a general practice. However, since these results are obtained in a cross-sectional design, no firm conclusions can be drawn. Nevertheless, it emphasises

the importance of a closer examination of the impact of the second year on the development of competence of trainees. As the trainees, at completion of their training, performed better than the volunteering (and thus highly motivated) GP-trainers, the implication of this finding for the teaching of clinical skills by GP-trainers needs also to be further investigated.

Chapter 5

As adequate communication between patient and physician is essential for good clinical practice, postgraduate training in general practice should contribute to a satisfactory acquisition of communication skills by trainees. From the literature it follows that training in communication skills that is provided through the years of training (longitudinal) and that is embedded in a rich clinical context (e.g. clinical clerkships, postgraduate training), is most effective. Since the postgraduate training under examination provided such conditions, the study described in this chapter investigated the effect of such a communication training on the acquisition of communication skills throughout the years of postgraduate training as well as whether a satisfactory level was achieved at the end of training. In a longitudinal design the communication skills of a national sample of trainees were assessed at the start and at the end of training. In order to measure communication skills, real life consultations were videotaped. The consultations were rated, using the MAAS-Global scoring list. The results were compared to the results of a reference group of experienced GPs.

The mean score slightly increased at the end of training (2.4) compared to the start (2.2). No significant difference was found between the final results of the trainees and the reference group. According to the criteria of the rating scale the performance of the trainees and GPs was unsatisfactory.

These findings suggest that communication skills do not improve in a three-year postgraduate training that comprised both a rich clinical context and a longitudinal training of communication skills, and that an unsatisfactory level still exists at the end of training. Moreover, GPs do not acquire communication skills during independent practice, as they perform comparable to the trainees. This all emphasises the need of further research into the measurement of communication skills, the teaching procedures and the role of the GP-trainer as a model.

Chapter 6

From a postgraduate training programme for general practice, it may be expected that it not only will support the transfer of actual knowledge and skills but also the extent to which trainees consider themselves competent (self-perceived clinical competence). This chapter reports a study into the effect of a three-year training programme that comprises both a rich clinical context and additional special educational programmes, on the development of self-perceived clinical competence. Moreover, because little is known about predictors of self-perceived clinical competence, the relation between the growth of self-perceived competence and several background variables that may potentially influence it (i.e. age, gender) was also explored. In a longitudinal design a national sample of the trainees completed a questionnaire, assessing self-perceived knowledge, clinical skills and consultation skills, at the start and at the end of training. Information about potentially influencing background variables (age, gender, prior medical experience, the effort someone has spent upon her/his education, insight in weak and strong areas of clinical competence and objectively measured knowledge and skills levels) was collected among them as well.

Growth of self-perceived clinical competence was found between the start and finish of the training programme and a satisfactory level was achieved at the end of the course. Self-perceived consultation skills increased more than self-perceived knowledge and clinical skills. The background variables did not relate in any way with the growth of self-perceived clinical competence.

These results suggest that a three-year postgraduate training in general practice satisfactory contributes to the feeling of being competent and that the feeling of being competent in consultation skills is more enhanced than that of knowledge and clinical skills. However, no explanation for the increase was found. Since self-assessment of competence is widely used as an assessment procedure in medical education and the importance of it for self-directed learning and professional autonomy is well established, more study into the theoretical concept of self-assessment and into the factors contributing to the feeling of being competent is required.

General conclusions and discussion

In this chapter the main findings are outlined and discussed, and recommendations for the future are given.

The methodological studies have provide evidence that (1) the Borderline Regression method is a credible, reliable and feasible standard setting procedure for an OSCE; and (2) a knowledge test of skills can be used as an instrument to estimate the level of clinical skills, especially for group evaluation.

With regard to the main research questions, four major conclusions can be drawn. Firstly, a three-year postgraduate training programme for general practice contributes satisfactory to the acquisition of knowledge and clinical skills, and to the self-perceived clinical competence of trainees; it does not seem to improve communication skills. Secondly, a three-year training prepares trainees better for clinical practice than a two-year training. Thirdly, the second block of the programme, in which trainees rotate through hospitals and the like, seems to contribute less to the acquisition of knowledge relevant to general practice and clinical skills than the first and third blocks of training in a general practice. Possibly, this also applies for communication skills. Finally, trainees, at the end of their training, perform better than (clinical skills) or comparable to (communication skills) experienced GPs.

In the methodological considerations the following issues are discussed. Firstly, the design of the study is considered. In this study the naturalistic environment of postgraduate training has been subject of investigation. Although this approach provides more relevant evidence than an experimental design, interpretations of the results are more difficult. To provide the best available evidence, the change of knowledge and skills across years of training has been examined. Secondly, in a study performed in ongoing curricula, unintended selection is more likely to occur than in an experimental, controlled study. Potential selection bias was evaluated by analysing differences between participating and missing trainees. These analyses revealed no statistically significant differences between both groups. Thirdly, the sample size of two sub-studies was rather small (47 and 25 participants). Although this size limits general conclusions, the fact that participants in both studies were representative for the entire group is reassuring.

In the discussion the following issues are considered. Firstly, this study focused on national evaluation. However, training institutes may differ in the way they administer the training programme and this may lead to differences in the acquisition of clinical competence. Therefore, a closer examination of differences between institutes is recommended. Secondly, this thesis focused on the outcome of postgraduate training. Although outcome measures are more directly related to the quality of the training programme, evaluation of the process of education is necessary and valuable as well. This especially applies for Dutch postgraduate training in general practice, since a national intended curriculum is supposed to be administered by eight different training institutes. As this study was part of a broader evaluation of postgraduate training (the EVA-H project), the process of education has been examined as well. However, this part of the evaluation has not been reported in this thesis. It provided information about the educational encounters between GP-trainees and GP-trainers, the education during clinical clerkships, the institutional education and the opinion of the GP-trainees about their training programme. For the future it is recommended to examine the outcome and the process of postgraduate training, to enable connections between both aspects. Thirdly, clinical competence was assessed according to the conceptual framework of Miller ('knows', 'knows how', 'shows how' and 'does') and assessment instruments were used in which competence is measured by quantitative, explicit and, as far as possible, objective criteria. This approach has difficulty in assessing attitude, the more professional aspects of clinical competence (e.g. the logic and suppleness with which the patient problem is applied) and the integration of the several aspects of clinical competence (knowledge, skills and attitude). Therefore, it is recommended to use qualitative measures, which rely on professional and patient judgement, as well. Moreover, only two of the four main tasks of the GP have been assessed: task 1, tasks related to the process of providing care and task 2, tasks related to certain categories of patients, complaints and diseases. These tasks refer to the more direct and traditional aspects of clinical competence in general practice. The two other tasks are related to the conditions under which the care is provided (task 3) and to the doctor as a person (task 4). These tasks refer to a broader approach of competence in general practice. It is recommended to assess this broader competence as well. Since adequate instruments to assess these skills do not exist, valid and reliable instruments have to

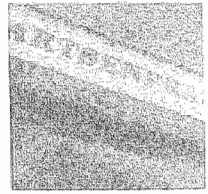


be developed. Fourthly, the influence of gender on learning and assessment has not been considered. From the literature we know that female and male learning styles differ and that gender has influence on the rating of performance. Therefore, it is recommended to consider gender in future evaluation studies. Finally, a (quasi-) experimental research approach was adopted. This approach can help to answer questions like 'how many' and 'how often'. Because learning and teaching involve many more dimensions, emotional, cognitive, attitudinal and physical, and because the process of learning is personal, subjective and highly variable, a qualitative research approach is required as well that may shed light on the questions of 'why' and 'how'. Therefore, it is recommended to use both approaches to research.

The chapter finishes with some practical recommendations for the near future, since a new curriculum, based on the principles of modular teaching and self-directed learning and with new learning outcomes, will be implemented in the next years. The following recommendations are given. Firstly, evaluation will provide the necessary evidence that the objectives of the new curriculum are indeed achieved. To perform such evaluation, a national centre for evaluation, bringing together qualified educationalists and GPs, remains an essential condition. Secondly, as the findings of the study indicated that the trainees acquired less knowledge and skills relevant to general practice during the clinical clerkships, it is questioned (1) to what extent the chosen learning environment and its position in the curriculum are appropriate to achieve the learning objectives of this period and (2) to what extent the assessment instruments were capable to measure the knowledge and skills acquired during this period. It will therefore be essential to redefine which knowledge, skills and attitude are necessary for a GP that cannot be learned in the learning environment of a general practice; which learning environment will be appropriate for this purpose; when and how this education has to take place; and how it can be assessed. Thirdly, the results of the study into communication skills are material for new questions to be answered: what are essential communication skills for general practice, how can they be learned and put into practice, what is the influence of the several parts of the curriculum on the acquisition of these skills and how can they be assessed?

Fourthly, the role of the experienced GP in the teaching of clinical and communication skills and as a learning model for trainees, needs further investigation. Fifthly, in this study standard setting procedures for an OSCE have been examined. Comparable studies are necessary for the knowledge tests. Finally, it will be interesting to learn more about self-assessment of clinical competence.





SAMENVATTING VAN DIT PROEFSCHRIFT

Dit proefschrift richt zich op de ontwikkeling van klinische competentie tijdens de huisartsopleiding. Na een inleiding, worden in zes hoofdstukken de resultaten van het onderzoek besproken, gevolgd door algemene conclusies en een discussie.

Inleiding

Omdat er steeds meer vraag is naar eerstelijnszorg en de voorziening ervan steeds complexer wordt, zijn er de afgelopen 30 jaar in verschillende landen opleidingen voor huisartsen ontwikkeld. Hoewel deze opleidingen er naar streven huisartsen in opleiding (haio's) goed voor te bereiden op hun toekomstige taken, is onvoldoende wetenschappelijk aangetoond of er ook bekwame huisartsen worden afgeleverd. Daarom is evaluatie-onderzoek nodig. Dit proefschrift beschrijft een dergelijk onderzoek naar de huisartsopleiding in Nederland.

Deze opleiding startte in 1973 en wordt door acht instituten, verbonden aan de afdelingen huisartsgeneeskunde van de acht universiteiten, verzorgd. Aanvankelijk duurde de opleiding een jaar. In 1985 werd de commissie 'Curriculum Constructie Beroepsopleiding tot Huisarts' (CCBOH) ingesteld die richtlijnen voor de structuur, de duur en de leerdoelen van het curriculum opstelde. Het Basistakenpakket voor de huisarts van de Landelijke Huisartsen Vereniging fungeerde als basis voor de inhoud van het onderwijs. Daarin worden vier hoofdtaken onderscheiden: (1) taken afgeleid van het hulpverleningsproces in de huisartsgeneeskunde; (2) taken afgeleid van bepaalde categorieën patiënten, klachten en ziekten; (3) ondersteunende taken (in staat zijn tot het voeren van een huisartspraktijk, het samenwerken met andere hulpverleners en het behouden van professionele competentie); en (4) taken die samenhangen met het persoonlijk functioneren van de huisarts. De acht opleidingsinstituten accepteerden de richtlijnen van de commissie CCBOH, zodat vanaf 1988 een grotere mate van overeenstemming bestaat tussen hun onderwijsprogramma's. Hoewel de commissie CCBOH een driejarige opleiding adviseerde, besloot de overheid tot een tweejarige opleiding in 1988. Verlenging naar drie jaar werd toegestaan in 1994, nadat was aangetoond dat niet alle inhoudelijke aspecten aan bod kwamen in de tweejarige opleiding. Sinds 1987 werken de acht opleidingsinstituten samen in de Stichting Verenigde Universitaire Huisartsopleidingen (SVUH) met als doel gezamenlijke activiteiten te ontplooiën ten aanzien van de

toetsing van haio's en de evaluatie van het onderwijs. In dit kader is een breed, landelijk evaluatie-onderzoek naar de huisartsopleiding uitgevoerd in de periode 1995-1999 (het EVA-H project).

Dit proefschrift beschrijft een deel van het EVA-H onderzoek en richt zich op het resultaat van de opleiding: (1) in welke mate beheersen haio's aan het einde van hun opleiding de taken van het Basistakenpakket? (2) is er een verschil tussen de tweejarige en de driejarige opleiding? en (3) hoe ontwikkelt de klinische competentie zich gedurende de opleiding? Het onderzoek beperkte zich tot de evaluatie van de 'kennis' en 'vaardigheden' betreffende de eerste twee hoofdtaken van het Basistakenpakket (taken afgeleid van het hulpverleningsproces en taken afgeleid van bepaalde categorieën patiënten, klachten en ziekten). Tevens werden twee methodologische onderzoeken uitgevoerd die nodig waren voor de beantwoording van de hoofdvragen.

De specifieke onderzoeksvragen waren:

1. Wat is de toename van kennis gedurende een tweejarige en een driejarige huisartsopleiding, en leidt een driejarige opleiding tot een hoger kennisniveau terwijl de structuur en leerdoelen dezelfde blijven? (Hoofdstuk 1)
2. Wat zijn de geloofwaardigheid en betrouwbaarheid van een rationele en een empirische methode om de zak/slaaggrens te bepalen voor een stationstoets, en wat is het effect van een 'reality check' op de geloofwaardigheid en betrouwbaarheid van de rationele procedure? (Hoofdstuk 2)
3. Wat is, in de huisartsopleiding, de voorspellende validiteit van een 'kennis over vaardigheden' toets voor de uitvoering van vaardigheden in een stationstoets? (Hoofdstuk 3)
4. Hoe ontwikkelen klinische vaardigheden zich tijdens de driejarige huisartsopleiding, en hebben haio's een voldoende niveau bereikt aan het einde van hun opleiding? (Hoofdstuk 4)
5. Hoe ontwikkelen communicatievaardigheden zich tijdens de driejarige huisartsopleiding; en hebben haio's een voldoende niveau bereikt aan het einde van hun opleiding? (Hoofdstuk 5)
6. Wat is de toename van zelf ervaren klinische competentie van haio's in een driejarige huisartsopleiding, en hoe kan deze toename worden verklaard? (Hoofdstuk 6)

Hoofdstuk 1

In dit hoofdstuk wordt onderzocht of een driejarige huisartsopleiding tot een hoger kennisniveau leidt dan een tweejarige opleiding en in welke mate het onderwijsprogramma bijdraagt aan de verwerving van kennis gedurende beide curricula.

Kennis werd getoetst door middel van voortgangstoetsing. Dit houdt in dat in elke toets de cognitieve einddoelen van het curriculum worden getoetst. Daarmee kan de groei van kennis ten opzichte van het einddoel worden gemeten. Om de huisartsgeneeskundige relevantie van de getoetste kennis te garanderen werd casuïstiek uit de huisartspraktijk gebruikt. In een retrospectief gemengd longitudinaal design werden de toetsresultaten van haio's van de tweejarige opleiding (1992-1996) vergeleken met die van de driejarige opleiding (1994-1999).

Uit de resultaten van het onderzoek komt naar voren dat verlenging van de tweejarige opleiding met een jaar, terwijl de structuur en leerdoelen hetzelfde blijven, tot een hoger kennisniveau leidt. De toename van kennis in beide curricula was aanzienlijk in vergelijking met de toename van kennis in het basiscurriculum. Zowel in het tweejarig als driejarig curriculum vond de meeste groei van kennis plaats in de eerste acht maanden en tegen het einde van de opleiding. Beide curricula toonden de grootste stagnatie in groei aan het einde van het tweede blok, waarin haio's stage liepen in ziekenhuizen, verpleeghuizen en GGZ-instellingen.

Omdat kennis ten grondslag ligt aan elke vorm van klinische competentie, stelt deze bevinding vragen bij het effect van het tweede blok op de ontwikkeling van de klinische competentie van haio's, en bij de doeltreffendheid van het onderwijsprogramma als geheel. Nader onderzoek naar deze aspecten is nodig voordat harde conclusies kunnen worden getrokken.

Hoofdstuk 2

Het onderzoek dat in dit hoofdstuk wordt beschreven betreft een methodologische studie naar procedures om een zak/slaaggrens (cesuur) vast te stellen. Deze studie was nodig om het tweede deel van onderzoeksvraag 4 (of een voldoende niveau van klinische vaardigheden was bereikt aan het einde van de opleiding, zie hoofdstuk 4) te beantwoorden. Om te kunnen vaststellen of een



voldoende niveau is bereikt moeten toetsresultaten vergeleken worden met een norm die aangeeft of iemand minimaal voldoende bekwaam is. Omdat het bepalen van een dergelijke norm nogal willekeurig is, is het van belang dat de totstandkoming ervan verdedigbaar is. Aspecten die bijdragen aan de verdedigbaarheid zijn de geloofwaardigheid en betrouwbaarheid van de norm. Deze studie richtte zich op twee absolute procedures om tot een norm te komen voor een stationstoets. Als rationele procedure is een gemodificeerde Angoff methode (alleen schriftelijke instructies, geen gelegenheid tot discussie en een aangepaste tweede beoordeling) met en zonder 'reality check' gebruikt. Voor de empirische procedure is een nieuwe methode, verwant aan de Borderline Group methode, toegepast: de Borderline Regression (BR) methode. De geloofwaardigheid, de betrouwbaarheid en het effect van een 'reality check' werden onderzocht voor beide procedures. In het kader van de studie werd een stationstoets afgenomen bij een landelijke steekproef van haio's in het laatste (derde) jaar van de opleiding en bij een referentiegroep van ervaren huisartsen. De betrouwbaarheid werd vastgesteld met behulp van de generaliseerbaarheidtheorie. De geloofwaardigheid werd beoordeeld door de slaagpercentages van haio's en ervaren huisartsen met elkaar te vergelijken en door het verband tussen de cesuurscore en de moeilijkheidsgraad van de toets te berekenen.

De cesuur score bedroeg 73.4% voor de Angoff methode zonder 'reality check' (Angoff I), 66.0% voor de Angoff methode met 'reality check' (Angoff II) en 57.6% voor de BR methode. De betrouwbaarheden (uitgedrukt als de 'root mean square errors') waren 2.1% voor Angoff I en II, en 0.6% voor de BR methode. De slaagpercentages van de haio's en huisartsen waren 19% en 9% voor Angoff I, 66% en 46% voor Angoff II en 95% en 80% voor de BR methode. De correlatie tussen de moeilijkheidsgraad van de toets en de cesuurscore was 0.69 voor Angoff I, 0.88 voor Angoff II en 0.86 voor de BR methode.

Op basis van deze resultaten lijkt de BR methode een geloofwaardigere en betrouwbaardere procedure om een norm te bepalen voor een stationstoets dan een gemodificeerde Angoff methode, zelfs als deze wordt toegepast met een 'reality check'. De algemene conclusie is dat met de BR methode een redelijke, verdedigbare en haalbare norm voor een stationstoets kan worden vastgesteld.



Hoofdstuk 3

Het onderzoek in dit hoofdstuk betreft de tweede methodologische studie, die nodig was om het eerste deel van onderzoeksvraag 4 te beantwoorden (zie hoofdstuk 4). Dit keer richtten we ons op de evaluatie van klinische vaardigheden. De stationstoets wordt algemeen gezien als een van de meest valide methoden om deze vaardigheden te beoordelen. Echter, omdat grootschalige toepassing van een stationstoets gecompliceerd en kostbaar is, kunnen instrumenten die minder veeleisend zijn een alternatief vormen. Daarom onderzochten we, in de huisartsopleiding, de voorspellende validiteit van een 'kennis over vaardigheden' toets voor de uitvoering van vaardigheden in een stationstoets. De stationstoets bestond uit stations waarin technische, procedurele vaardigheden werden getoetst en stations waarin de klinische vaardigheden van een geheel consult werden beoordeeld (integrale stations). Drie hypothesen werden getoetst, betrekking hebbend op (1) de inhoud van de kennistoets (algemene kennis versus kennis over vaardigheden), (2) het beoordelingssysteem van de stationstoets (globaal versus scoringslijst) en (3) de inhoud van de stationstoets (technische stations versus integrale stations). Een 'kennis over vaardigheden' toets, een algemene kennistoets en een stationstoets werden afgenomen bij een landelijke steekproef van haio's van de driejarige huisartsopleiding. De correlaties tussen de verschillende toetsresultaten werden berekend.

Twee hypothesen werden bevestigd. Ten eerste, de 'kennis over vaardigheden' toets toonde meer verband met de stationstoets dan de algemene kennistoets (hypothese 1). Dit was vooral het geval voor de technische stations. Ten tweede, de 'kennis over vaardigheden' toets toonde meer verband met de technische stations dan met de integrale stations (hypothese 3). Tot slot, de 'kennis over vaardigheden' toets toonde meer verband met de integrale stations wanneer deze beoordeeld werden met een scoringslijst dan wanneer deze globaal beoordeeld werden (hypothese 2). Echter, voor de technische stations werd een dergelijk verschil niet gevonden.

Deze resultaten ondersteunen de voorspellende waarde van de 'kennis over vaardigheden' toets. Dit betekent dat, in de huisartsopleiding, een 'kennis over vaardigheden' toets gebruikt kan worden als instrument om het niveau van klinisch vaardigheden te schatten. Dit geldt vooral voor de evaluatie van groepen zoals in studies waarin het effect van een vaardigheidentraining wordt onderzocht of als screenings-

instrument bij het vaststellen van een cursusaanbod.

Deze schatting is preciezer naarmate de inhoud van de toets meer overeenkomt met de vaardigheden die onderzocht worden.

De resultaten betekenen echter niet dat een 'kennis over vaardigheden' toets de toetsing van vaardigheden met behulp van een stationstoets kan vervangen. Directe observatie van hoe een vaardigheid wordt uitgevoerd blijft meer valide en het stimuleert haio's tot oefening.

Hoofdstuk 4

Omdat klinische vaardigheden (waaronder wordt verstaan: anamnese, lichamelijk onderzoek, diagnose en beleid) een essentieel onderdeel zijn van het dagelijks werk van de huisarts, is het van groot belang dat een huisartsopleiding er toe bijdraagt dat haio's deze vaardigheden in voldoende mate verwerven. Daarom wordt in dit hoofdstuk onderzocht hoe klinische vaardigheden zich ontwikkelen gedurende de opleiding, en of een voldoende niveau is bereikt aan het einde ervan. Het betreft een onderzoek naar een driejarige huisartsopleiding waarin de verwerving van klinische vaardigheden grote aandacht kreeg. In een cross-sectioneel design werd het vaardighedeniveau door de jaren heen gemeten met behulp van een 'kennis over vaardigheden' toets, die de vorm van een voortgangstoets had. De toets werd afgenomen bij alle haio's in alle drie jaren van de opleiding. Om het eindniveau vast te stellen werd bij een landelijke steekproef van haio's in de laatste fase van hun opleiding een 'kennis over vaardigheden' toets en een stationstoets afgenomen. De stationstoets bestond uit stations waarin technische, procedurele vaardigheden werden getoetst en stations waarin de klinische vaardigheden van een geheel consult werden beoordeeld (integrale stations). De resultaten van beide toetsen werden vergeleken met een zak/slaag norm en met de resultaten van een referentiegroep van huisartsopleiders. De toename van klinische vaardigheden was 18% (van 30% naar 48%). De meeste groei vond plaats in de eerste zes maanden van de opleiding. Daarna werd een geleidelijke maar minder snelle toename gezien, behalve aan het einde van het tweede jaar waarin een stagnatie in groei werd waargenomen. Dit was het moment waarop haio's hun stageperiode in ziekenhuizen, verpleeghuizen en GGZ-instellingen afrondden. De haio's presteerden beter dan de huisartsopleiders, zowel op de kennistoets als op de stationstoets.



Ongeveer 90% van hen zou geslaagd zijn voor de kennistoets en 95% voor de stationstoets. De resultaten op de stationstoets tonen een opvallend onderscheid tussen de technische en integrale stations. Er zouden meer haio's geslaagd zijn voor de integrale stations dan voor de technische stations. Dit wijst er op dat integrale vaardigheden beter werden beheerst dan technische vaardigheden. Bovendien waren de resultaten van de integrale stations min of meer hetzelfde voor haio's en huisartsen, terwijl de huisartsen op de technische stations minder goed scoorden dan de haio's.

De bevindingen van deze studie geven aan dat een driejarige huisartsopleiding, waarin de verwerving van klinische vaardigheden een belangrijk doel is, bijdraagt aan een voldoende toename van deze vaardigheden. Integrale vaardigheden betrekking hebbend op gehele consulten, lijken beter verworven te worden dan technische, procedurele vaardigheden. De bevindingen wijzen er ook op dat het tweede jaar van de opleiding, waarin haio's stage lopen in ziekenhuizen e.d., minder bijdraagt aan het verwerven van klinische vaardigheden dan het eerste en derde jaar van de opleiding in een huisartspraktijk. Omdat de resultaten verkregen zijn uit een cross-sectionele studie, zijn hieruit echter geen harde conclusies te trekken. Desalniettemin, het geeft wel het belang aan van nader onderzoek naar het effect van het tweede jaar op de competentie ontwikkeling. Omdat de haio's aan het einde van hun opleiding beter presteerden dan de vrijwillig deelnemende (en dus zeer gemotiveerde) huisartsopleiders, dient ook de betekenis van deze bevinding voor het trainen van klinische vaardigheden door huisartsopleiders verder onderzocht te worden.

Hoofdstuk 5

Omdat adequate gespreksvoering tussen patiënt en arts essentieel is voor goed klinisch handelen, dient een huisartsopleiding er toe bij te dragen dat haio's voldoende communicatievaardigheden verwerven. Uit de literatuur weten we dat een communicatietraining die gedurende de gehele opleiding gegeven wordt (longitudinale training) en die plaats vindt in een rijke klinische context (bijv. co-assistentenschappen, huisartsopleiding), het meest effectief is. De communicatietraining van de studie waar dit hoofdstuk verslag van doet, voldoet aan deze kenmerken. Daarom is in deze studie onderzocht wat het effect van een dergelijke training was op de

verwerving van communicatievaardigheden tijdens de opleiding, en of een voldoende niveau was bereikt aan het einde ervan. In een longitudinaal design werden de communicatievaardigheden van een landelijke streekproef van haio's zowel aan het begin als aan het eind van de opleiding getoetst. De communicatievaardigheden werden beoordeeld aan de hand van consulten uit de dagelijkse praktijk die op video waren gezet. De consulten werden beoordeeld met behulp van de Maas-Globaal scoringslijst. De resultaten werden vergeleken met de resultaten van een referentiegroep van ervaren huisartsen. De gemiddelde score nam iets toe aan het einde van de opleiding (van 2.2 naar 2.4). Er werd geen significant verschil gevonden tussen de eindresultaten van de haio's en die van de referentiegroep. Volgens de criteria van de scoringslijst presteerden haio's en huisartsen onvoldoende.

Deze bevindingen suggereren dat communicatievaardigheden niet verbeteren in een driejarige huisartsopleiding met een longitudinale communicatie training die plaats vindt in een rijke klinische context, en dat aan het einde van de opleiding het niveau nog steeds onvoldoende is. Omdat het niveau van ervaren huisartsen vergelijkbaar is met dat van haio's, lijkt het dat huisartsen ook tijdens hun praktijktoefening geen communicatievaardigheden verwerven. Dit alles benadrukt de behoefte aan meer onderzoek naar het meten van communicatievaardigheden, de onderwijsmethodes en de rol van de opleider als leermodel.



Hoofdstuk 6

Van een huisartsopleiding mag men verwachten dat het niet alleen de overdracht van kennis en vaardigheden ondersteunt, maar ook de mate waarin haio's zichzelf competent achten (zelf ervaren competentie). In dit hoofdstuk wordt een onderzoek beschreven naar het effect van een driejarige huisartsopleiding op de ontwikkeling van zelf ervaren competentie. De onderzochte opleiding bood zowel onderwijs in een rijke klinische context (stages in huisartspraktijken, ziekenhuizen e.d.) als aanvullende speciale onderwijsprogramma's. Omdat er weinig bekend is over welke factoren bijdragen aan het gevoel van eigen competentie, is ook het verband tussen de groei van zelf ervaren competentie en verschillende achtergrondvariabelen die er mogelijk op van invloed zijn (bijv. leeftijd, geslacht) onderzocht. In een longitudinaal design heeft een landelijke steekproef van haio's

zowel aan het begin als aan het einde van de opleiding een vragenlijst ingevuld over zelf ervaren kennis, klinische en consultvoeringsvaardigheden. Hen werd ook gevraagd informatie te geven over mogelijk van invloed zijnde achtergrondvariabelen (leeftijd, geslacht, voorafgaande medische ervaring, de mate waarin iemand zijn best doet in de opleiding, inzicht in zwakke en sterke gebieden van klinische competentie en objectief gemeten kennis- en vaardighedenniveaus). Er werd een toename van zelf ervaren klinische competentie gevonden tussen begin en einde van de opleiding en aan het einde van de opleiding was de zelf ervaren competentie voldoende. Zelf ervaren competentie in consultvoeringsvaardigheden nam meer toe dan die van kennis en klinische vaardigheden. De achtergrondvariabelen toonden geen enkel verband met de groei van zelf ervaren klinische competentie.

De resultaten suggereren dat een driejarige huisartsopleiding voldoende bijdraagt aan het eigen gevoel van klinische competentie, waarbij het gevoel van competentie in consultvoering meer lijkt te worden gestimuleerd dan dat van kennis en klinische vaardigheden. Er werd echter geen verklaring gevonden voor deze toename. Omdat zelfevaluatie van competentie op grote schaal wordt gebruikt als evaluatiemethode in het medisch onderwijs en het belang ervan voor zelf gestuurd leren en professionele autonomie alom erkend wordt, is meer onderzoek naar het theoretische concept van zelfevaluatie en de factoren die bijdragen aan het eigen gevoel van competentie noodzakelijk.

Algemene conclusies en discussie

In dit hoofdstuk worden de belangrijkste bevindingen beschreven en besproken. Tevens worden aanbevelingen gedaan voor de toekomst.

De methodologische studies hebben aangetoond dat (1) de Borderline Regression methode een geloofwaardige, betrouwbare en haalbare procedure is om tot een zak/slaag grens te komen voor een stationstoets en (2) een 'kennis over vaardigheden' toets gebruikt kan worden als instrument om het klinisch vaardighedenniveau te schatten, vooral in geval van groepsevaluatie.

Ten aanzien van de centrale onderzoeksvragen kunnen vier belangrijke conclusies worden getrokken. Ten eerste, een driejarige huisartsopleiding draagt voldoende bij aan de verwerving van kennis



en klinische vaardigheden en de mate waarin haio's zichzelf competent achten; communicatievaardigheden lijken niet toe te nemen. Ten tweede, een driejarige opleiding bereidt haio's beter voor op de huisartsgeneeskundige praktijk dan een tweejarige opleiding. Ten derde, het tweede blok van de opleiding (de externe leerwerkperiode), waarin haio's stages lopen in ziekenhuizen e.d., lijkt minder bij te dragen aan de verwerving van huisartsgeneeskundig relevante kennis en klinische vaardigheden dan het eerste en derde blok met stages in een huisartspraktijk. Mogelijk geldt dit ook voor communicatievaardigheden. Tot slot, haio's die de opleiding bijna voltooid hebben, presteren beter dan (klinische vaardigheden) of vergelijkbaar met (communicatievaardigheden) ervaren huisartsen.

In de methodologische beschouwingen zijn de volgende problemen aan de orde gekomen. Ten eerste is het design van de studie besproken. In deze studie is de natuurlijke omgeving van de huisartsopleiding onderwerp van onderzoek geweest. Hoewel deze benadering relevanter bewijs levert dan een experimenteel design, is het moeilijker om de resultaten te interpreteren. Om het best mogelijke bewijs te leveren is de verandering van kennis en vaardigheden over de opleidingsjaren heen bestudeerd. Ten tweede is in een onderzoek dat uitgevoerd wordt in lopende curricula, de kans groter dat er onbedoelde selectie optreedt dan in een experimentele, gecontroleerde studie. Door verschillen tussen deelnemende en uitvallende haio's te analyseren is mogelijke selectie bias onderzocht. Deze analyses toonden geen statistisch significante verschillen tussen beide groepen. Ten derde was de steekproef van twee deelonderzoeken nogal klein (47 en 25 deelnemers). Hoewel deze aantallen het trekken van algemene conclusies beperken, is het feit dat de deelnemers in beide studies representatief waren voor de gehele groep geruststellend.

In de discussie zijn de volgende onderwerpen besproken. Ten eerste, deze studie richtte zich op landelijke evaluatie. Echter, opleidingsinstituten kunnen verschillen in de wijze waarop het onderwijsprogramma wordt uitgevoerd en dit kan leiden tot verschillen in verwerving van kennis en vaardigheden. Daarom wordt een nader onderzoek naar verschillen tussen instituten aanbevolen. Ten tweede, dit onderzoek richtte zich op het resultaat van de opleiding. Hoewel uitkomstmaten meer rechtstreeks gerelateerd zijn aan de kwaliteit van



het onderwijsprogramma, is evaluatie van het onderwijsproces eveneens nodig en waardevol. Dit geldt vooral voor de Nederlandse huisartsopleiding waarvan verondersteld wordt dat een landelijk vastgesteld curriculum uitgevoerd wordt door acht verschillende opleidingsinstituten. Omdat deze studie onderdeel was van een bredere evaluatie van de opleiding (het EVA-H project), is het onderwijsproces wel onderzocht maar niet verslagen in dit proefschrift. Dit onderzoek gaf informatie over de leergesprekken tussen haio's en huisartsopleiders, het onderwijs tijdens de externe leerwerkperiode, het instituutsonderwijs en de mening van haio's over het onderwijsprogramma. Voor de toekomst wordt aanbevolen om zowel proces als resultaat van de opleiding te onderzoeken, zodat verbanden tussen beide kunnen worden gelegd. Ten derde, klinische competentie is getoetst volgens het conceptuele schema van Miller ('weet', 'weet hoe', 'toont hoe' en 'doet') en er is gebruik gemaakt van instrumenten die competentie meten met behulp van kwantitatieve, expliciete en, voor zover mogelijk, objectieve criteria. Door deze benadering zijn attitude, de meer professionele aspecten van klinische competentie (bijv. de logica en soepelheid waarmee een patiëntenprobleem wordt benaderd) en de integratie van de verschillende aspecten van klinische competentie (kennis, vaardigheden en attitude) moeilijker te beoordelen. Het wordt daarom aanbevolen om ook gebruik te maken van kwalitatieve maten die berusten op professionele en patiëntenoordelen. Bovendien zijn slechts twee van de vier hoofdtaken van de huisarts geëvalueerd: taak 1, taken betrekking hebbend op het hulpverleningsproces, en taak 2, taken betreffende bepaalde categorieën van patiënten, aandoeningen en ziektes. Deze taken verwijzen naar de meer directe en traditionele aspecten van klinische competentie. De twee andere taken richten zich op de condities waaronder de zorg wordt verleend (taak 3) en op de dokter als persoon (taak 4). Deze taken verwijzen naar een bredere benadering van huisartsgeneeskundige competentie. Aanbevolen wordt om deze bredere competentie ook te evalueren. Omdat adequate instrumenten om dit te toetsen ontbreken, zullen betrouwbare en valide instrumenten ontwikkeld moeten worden. Ten vierde, er is geen rekening gehouden met de mogelijke invloed die 'gender' heeft op leren en toetsing. Uit de literatuur weten we dat de leerstijlen van mannen en vrouwen verschillen en dat 'gender' invloed heeft op de beoordeling van studenten bij de uitvoering van vaardigheden. Het wordt daarom aanbevolen om 'gender' te

betrekken bij toekomstige evaluaties. Tot slot, er is gekozen voor een (quasi) experimentele benadering van onderzoek. Deze benadering kan van nut zijn bij de beantwoording op vragen als 'hoeveel' en 'hoe vaak'. Omdat leren en toetsing veel meer dimensies omvatten, zoals emotionele, cognitieve, houdings- en fysieke, en omdat het leerproces persoonlijk, subjectief en zeer variabel is, is een kwalitatieve benadering van onderzoek eveneens nodig om antwoord te geven op de vragen 'waarom' en 'hoe'. Daarom worden beide onderzoeksbenaderingen aanbevolen.

Het hoofdstuk eindigt met enkele praktische aanbevelingen voor de nabije toekomst, omdat een nieuw curriculum, gebaseerd op de uitgangspunten van modulair onderwijs en zelf gestuurd leren en met nieuwe eindtermen, in de komende jaren geïmplementeerd zal worden. De volgende aanbevelingen worden gedaan. Ten eerste, evaluatie zal het noodzakelijke bewijs leveren dat de doelen van het nieuwe curriculum daadwerkelijk worden gehaald. Om een dergelijke evaluatie uit te voeren is een nationaal evaluatiecentrum, waarin onderwijskundigen en huisartsen samenwerken, een essentiële voorwaarde. Ten tweede, omdat de resultaten van deze studie aangeven dat de haio's minder huisartsgeneeskundig relevante kennis en vaardigheden verwierven in de externe leerwerkperiode, kan men zich afvragen (1) in hoeverre deze leeromgeving en de plaats ervan in het curriculum geschikt waren om de leerdoelen van deze periode te bereiken en (2) in welke mate de toetsen in staat waren om de kennis en vaardigheden, die in die periode verworven werden, te meten. Het zal daarom nodig zijn om opnieuw te bepalen welke kennis, vaardigheden en attitude noodzakelijk zijn voor de huisarts die niet geleerd kunnen worden in de leeromgeving van een huisartspraktijk; welke leeromgeving geschikt is voor dit doel; wanneer en hoe dit onderwijs moet plaatsvinden; en hoe het kan worden getoetst. Ten derde, de resultaten van het onderzoek naar communicatievaardigheden roepen nieuwe vragen op die een antwoord behoeven: wat zijn essentiële communicatievaardigheden voor de huisarts, hoe kunnen ze worden aangeleerd en in praktijk gebracht, wat is de invloed van de verschillende onderdelen van het curriculum op de verwerving van deze vaardigheden en hoe kunnen ze worden getoetst? Ten vierde, de rol van de ervaren huisarts als opleider in klinische en communicatievaardigheden en als leermodel voor haio's behoeft nader onderzoek.



Ten vijfde, in deze studie zijn procedures om een zak-/slaaggrens vast te stellen voor een stationstoets onderzocht. Vergelijkbare studies zijn nodig voor de kennistoetsen. Tot slot, het zou interessant zijn om meer te weten te komen over de zelfevaluatie van klinische competentie.

DANKWOORD

Nu de klus grotendeels is geklaard, is het moment van bedanken aangebroken. Waren er aanvankelijk veel mensen bij het onderzoek betrokken, nu is alleen de promovenda nog aan het werk. Dit dankwoord is dan ook een klein retrospectief onderzoek naar wie er allemaal toe hebben bijgedragen dat dit proefschrift tot stand is gekomen. Aan het onderzoek (het EVA-H project), waar dit proefschrift verslag van doet, deden acht opleidingsinstituten mee. Er is gebruik gemaakt van een groot aantal toetsen en vragenlijsten. Daarom is het aantal betrokkenen heel groot. Ik zal ongetwijfeld mensen vergeten. Mijn dank geldt ook hen!

Lisa Tan, jij bent nummer een. Dankzij jouw grote standvastigheid en doorzettingsvermogen is het EVA-H project van de grond gekomen en uitgevoerd. In eerste instantie slaagde je er niet in om mij te strikken voor het onderzoek. Het lot gaf ons een tweede kans; en dit keer had je meer succes. Ik stapte achter op je EVA-H fiets en samen hebben we het project afgerond. Jij werd co-promotor bij de wetenschappelijke verantwoording van het onderzoek. Al ging je met pensioen in 2000, je bleef naast me staan zowel inhoudelijk als persoonlijk, toen de omstandigheden waarin ik aan mijn proefschrift moest werken wel erg moeilijk werden. Je geloof in mij is een grote steun geweest. Ik heb veel geleerd van de manier waarop jij wetenschap bedrijft: precies en kritisch, met grote distantie én betrokkenheid, en met veel oog voor het detail maar ook met een groot relativeringsvermogen.

De huisartsen in opleiding die bereid waren om gedurende de drie jaren van hun opleiding vragenlijsten in te vullen en zich te onderwerpen aan allerlei toetsen: hun gegevens vormen de basis van dit onderzoek. Mede dankzij hun trouw is dit onderzoek een succes geworden. Onontbeerlijk daarbij was de steun van hun huisartsopleiders en docenten.

Herman Düsman en Paul Ket, jullie hadden een zeer belangrijk aandeel in het ontwerp en de uitvoering van het EVA-H project. Dat jullie je werk goed hebben gedaan, blijkt uit de waardevolle artikelen die er uit zijn voortgekomen. Herman, jij hebt ook bijgedragen aan de wetenschappelijke verantwoording van het project door je kritische en degelijke methodologische inbreng, en je hulp bij de data-analyses. Onze wetenschappelijke discussies in kamer 408b waren een bron van inspiratie.

Cees van der Vleuten en Richard Grol. Promotor betekent iemand die naar voren beweegt: jullie hebben het mij mogelijk gemaakt - ieder op eigen wijze - vooruit te komen in de wetenschap. Niet alleen door mij te leren hoe je onderzoek doet, maar vooral ook hoe je er verslag van doet. Cees, jouw enthousiasme voor bevindingen uit het onderzoek, je voortdurende bereidheid om mijn vragen te bespreken en je deskundigheid in onderzoek van onderwijs, waren stimulerend en steunend. Richard, jouw bijdrage zat hem juist in de afstand, die deels voortkwam uit je positie als tweede promotor en deels uit het feit dat jouw deskundigheid meer bij de kwaliteit van de huisartsgeneeskunde dan bij medisch onderwijs ligt. Daarom leverde je steeds een kritische, 'helikopter' bijdrage aan het onderzoek. Hoewel ik het niet altijd even gemakkelijk vond, zette je aan tot verdieping en verbreding.

De leden van de werkgroep EVA-H, Luc van Berkestijn, Peter Boendermaker, Gerrit Locher, Bas Maiburg, Richard Starmans, Han Stolk, Myrra Vernooij-Dassen en Ernst van der Wiele, jullie vormden de verbinding tussen 'het bureau' en de opleidingsinstituten. Dankzij jullie inzet voor het project werden de plannen omgezet in daden.



Koos Jansen, rustige rots in de branding. Je betrokkenheid bij het EVA-H project begon bij de stationstoets. Terwijl je druk bezig was met je eigen promotie, verbond je je aan deze grote operatie. Ik was er nauw bij betrokken en weet hoeveel werk dit is geweest. Het geeft aan hoe betrokken je bent bij de training en toetsing van klinische vaardigheden, en bij bureau TEO. Daarna wilde je me graag ondersteunen bij mijn promotie. Samen met Lisa en Herman waren jullie mijn 'interne promotieteam', waar ik in alle rust mijn vragen, ideeën en problemen kon bespreken. Jouw inbreng was kritisch, ondersteunend en stimulerend. Je deed me steeds weer in mezelf geloven als het even niet mee zat. Je bent een creatief denker en nooit te beroerd om naar betere formuleringen te zoeken.


De leden van de werkgroep vaardigheden en de trainers en beoordelaars van de stationstoets, in het bijzonder Jacqueline Bloemen, Peter Boendermaker, Jaap Buis, Peter van Hasselt, Bas Maiburg en Charles Verhoeff: dankzij jullie inspanningen werd de stationstoets een succes. Jacqueline, jij leverde tevens een belangrijk aandeel in de 'kennis-over-vaardigheden' toetsen.

Peter Zuithoff, jij verzorgde de data van EVA-H, en dat zijn er heel veel. Steeds was je bereid mijn verzoeken om analyses uit te voeren, zonder protest of reserve. Daarnaast was je nauw bij de inhoud betrokken, je dacht en deed mee. Je relativiseringsvermogen blijkt onder andere uit de 'grappige e-mails' die je regelmatig toestuurt met daarin een kritische kijk op de gebeurtenissen in de wereld.

De leden van de werkgroep kennis en de medewerkers van bureau TEO die in de loop der jaren hebben bijgedragen aan de productie van de kennistoetsen. Dankzij jullie inspanningen kwam er aanvankelijk elke vier maanden en sinds 1996 elke zes maanden een nieuwe toets.

Paul Ram, MAAS-Globaal specialist. Jij was bereid de observatoren te trainen voor de beoordeling van de communicatievaardigheden van de huisartsen in opleiding.

De observatoren van de videobanden: dank voor jullie medewerking.



De aanstuurders, bestendigers, ondersteuners, begeleiders en toezichthouders van het EVA-H project. Vic Tielens, jij was inspirator. De HVRC drong aan op landelijke evaluatie. De hoofden van de acht opleidingsinstituten, verenigd in de SVUH, gingen akkoord met het projectvoorstel en ondersteunden het project tot en met dit proefschrift. De SBOH gaf subsidie. Toen het project eenmaal liep, zorgde het DB van de SVUH voor de voortgang. Aanvankelijk waren dat Vic Tielens, Vic Dubois en Frank Almekinders. Later waren dat Piet Schoonheim en Ron Pieters. Jullie zorgden er ook voor dat de wetenschappelijke verantwoording van het project, nadat het project zelf was afgerond, doorgang kon vinden. De Commissie van Toezicht, in het bijzonder Lourens Kooij en Ruud Hogenhuis, bewaakte de voortgang van het project op afstand. De Wetenschappelijke Commissie, tot slot, keek toe of het project voldoende wetenschappelijk werd uitgevoerd.

Bureau TEO heeft mij de ruimte gegeven dit onderzoek te kunnen doen. Een paar mensen wil ik speciaal noemen. Elbrig en Paula (en Hans Eeckhout, toen je nog bij ons werkte), jullie vormen de basis van het bureau, met al die in- en uitvliegende parttimers. Jullie aandeel in de productie en verwerking van de toetsen en vragenlijsten was groot. Tom Segers, het onderzoek was al bijna klaar toen jij bij bureau TEO kwam. Toch heb je een essentiële rol gespeeld in de afronding ervan

door er voor te zorgen dat het proefschrift een zachte landing kreeg. Je stuurkunst is voortreffelijk. Wilma Spinnewijn, jij was de afgelopen jaren eindverantwoordelijke voor de kennistoets. Ik heb groot respect voor je volharding, het is je steeds weer gelukt de toets op tijd klaar te hebben.

Huisartsenpraktijk Anthoniedijk, jullie gaven mij de gelegenheid om me gedurende twee en een half jaar geheel en al 'aan de wetenschap' te wijden. Na 15 jaar dokteren in een achterstandswijk werd deze periode een oase van rust. Dat ik ben teruggekomen ligt deels aan jullie. De gemoedelijkheid, solidariteit en betrokkenheid waarmee wij met elkaar werken, kan en wil ik niet missen. Ook heeft deze periode mij duidelijk gemaakt dat ik het werk als huisarts met zo'n gemêleerde en boeiende patiëntenpopulatie veel te leuk vind om er mee te stoppen. Ik gun iedere dokter zo'n periode van afstand; het geeft je de ruimte om te beseffen hoe leuk het is om huisarts te zijn. Niek, maatje van het eerste uur. Jij bent de motor van de praktijk. Met je rustige daadkracht regel je veel en zorg je voor stabiliteit. Ik ben je dankbaar voor je solidariteit met mij door de jaren heen. Onze wegen kruisen elkaar niet alleen in de praktijk, maar ook in onze betrokkenheid bij de huisartsopleiding: je bent huisartsopleider en voorzitter van de werkgroep kennis. Jij moest wel paranimf worden; en gelukkig wilde je dat ook. Natascha, super-doktersassistente / bijna-huisarts. Jouw zorg, betrokkenheid, trouw en deskundigheid zijn van onschatbare waarde. Ik vrees dat je echt onmisbaar bent. Dank voor de vele jaren van echte samenwerking. Roeland, dankzij je besliste optreden zonder veel woorden zorg jij ook voor rust en stabiliteit in de praktijk. Dank voor de ruimte die je me gaf dit onderzoek te doen. Licky, jij zou me de eerste maanden van mijn verlof vervangen, je wilde je niet vastleggen voor daarna. Gelukkig ben je gebleven zodat ik mijn onderzoek in alle rust kon doen. Het was erg prettig te weten dat de patiënten bij jou in veilige handen waren. Wij delen het enthousiasme voor 'het kleurrijk palet' aan patiënten in de praktijk. Dat je er nog steeds bent en hopelijk nog lang zult blijven, is een zeer prettige 'bijkomstigheid'.

Mijn familie. Lieve ouders, Marie-Thérèse en Mebius, ik heb het proefschrift niet voor niets aan jullie opgedragen. Niet alleen door de bagage die ik van jullie heb meegekregen, maar ook door de steun en het goede voorbeeld die jullie me nog steeds geven, ben ik in staat veel te verzetten, waaronder het schrijven van een proefschrift.

Het voorbeeld van een oma die tot haar dood op zeer hoge leeftijd van het leven genoot en er wat van maakte, komt daar nog bij. Dank ook voor de rustige werkkamer met computer (!), de vele kopjes koffie en thee, en de lekkere lunches tijdens de afronding van het proefschrift. Mebius, jouw inzet voor medisch onderwijs hebben er vast toe bijgedragen dat ik op dit onderwerp promoveer. IJsbrand, ik vroeg je om me te helpen bij het Engels. Dat jij je daartoe niet beperkte, had ik kunnen vermoeden. Ongevraagd werd je mijn eerste externe beoordelaar. Met je ruime internationale wetenschappelijke ervaring en je kritische geest, gaf je pittig commentaar op mijn concepten. Dat ontaardde soms in leuke wetenschappelijke discussies. Met je grondige beoordeling van mijn teksten, moest ook jij wel paranimf worden. Maarten, ik vroeg jou om de voorkant van het proefschrift te maken. Dat deed ik, omdat ik vind dat je met veel zorg en een scherp oog de prachtigste dingen maakt. Met je ontwerp heb jij mijn oordeel wederom bevestigd.



Lieve Boy en Deborah, jullie zorgden ervoor dat ik niet te lang en niet te vaak doorwerkte. Daarmee hebben jullie een hele belangrijke bijdrage geleverd. Het is namelijk heel goed om regelmatig iets anders te doen als je 'aan je boek werkt' - zoals jullie het noemden -. Dan kan je er daarna weer beter over nadenken. Jullie vonden het spannend dat ik een boek schreef. Alleen de inhoud kon jullie niet zo interesseren. Boy, jij vroeg me om als volgend project een kinderboek te schrijven. Want dit boek, al die lappen tekst en dan ook nog in het Engels, brrr, niets aan eigenlijk.

Lieve Carol, je onwankelbaar vertrouwen in mij, je vaste overtuiging dat ik dit onderzoek moest doen en je plezier in mijn wetenschappelijke successen, hebben mij enorm gesteund. Hoewel je zegt dat je er in praktische zin niet veel van hebt gemerkt, heb je toch stilletjes en stiekempjes, zeker in de laatste maanden, taken van me overgenomen. De discussies over wetenschap die wij al jaren voeren en jouw wetenschapsopvatting en -praktijk, hebben mij wetenschappelijk gevormd. Dit alles en nog veel meer maakt jou tot mijn nummer een.

CURRICULUM VITAE

Anneke Kramer was born on July 13th, 1955 in Utrecht, The Netherlands. She finished her schooling (Gymnasium-B) in 1973 at the Stedelijk Gymnasium in Utrecht. She entered Medical School at Utrecht University. She participated in several educational committees within the faculty. After acquiring her medical degree in January 1981, she was involved in the organisation of a part-time postgraduate training in general practice. She also worked as a practitioner in an infant welfare centre. From November 1981 to November 1983 she followed the first part-time postgraduate training, at the University of Utrecht. Her GP-trainer was Henk Thiadens in Amersfoort. In 1984 she started a general practice in an unprivileged district of the city of Utrecht, in association with Niek de Grunt. From 1985 to 1987 she worked as a part-time teacher at the Medical School of the Free University in Amsterdam. Since 1987 she has been a part-time staff member at the National Centre of Evaluation of Postgraduate Training in General Practice. Her main tasks are the development and production of knowledge and skills tests. In 1999 she started with the study presented in this thesis.